PULSE FREQUENCY TECHNIQUES FOR AUTOMATIC CONTROL

by

Terence Haydn Thomas B.A.

A thesis submitted to the

UNIVERSITY OF WARWICK

for the degree of Doctor of Philosophy

September 1971

STATEMENT

The material contained in this thesis is the result of my own research, except where specifically indicated to the contrary. It has not formed part of a submission for a degree at any university.

I wish to thank the School of Engineering Science at the University of Warwick for the provision of facilities, instruction and a congenial working atmosphere, Serck Ltd. for financial support over 1 year and Courtaulds Educational Trust for such support over 3 years. I also thank my supervisors Professor J.L. Douce and Dr. M.T.G. Hughes. Among those who have helped me with advice, I wish to particularly mention Dr. J. Monk and Dr. J.V. Comfort of this university and Mr. J.D. Martin of the University of Bath.

T.H. Thomas

T.H. Thomas

COVENTRY, 30.9.1971

ABSTRACT

This thesis describes an investigation into the suitability of pulse frequency modulation (PFM) as a standard form of signal for representing quantities in process control. The encoding and decoding of PFM signals into both analogue and digital forms is examined in some detail. PFM is shown to be well suited for high accuracy telemetry at moderate cost, provided ample channel bandwidth is available. The processing of the information in PFM signals by means of binary logic devices is treated systematically. Functional building blocks are identified, and shown to be capable of performing all the basic algebraic and differential operations needed for control.

The thesis concludes with an examination of applications and a discussion of PFM transducers, actuators and heirarchical control schemes. The performance constraints of two different process controllers are identified. Both controllers show 'P + I' action; one works continuously, the other has a cyclic action; both employ PFM techniques. They are shown to offer dynamic responses similar to those of conventional analogue controllers, in conjunction with high accuracy (e.g. errors less than ½%), computer compatibility and the facility for digital display.

LIST OF CONTENTS

Bound in papers :

'A rapid pulse-frequency to number converter', Electron. Lett., 7, 12, 17th June 1971, pp 345-6.

'Performance of a digital two-term controller', Automatica, 5, 1, Jan. 1969, pp 67-75.

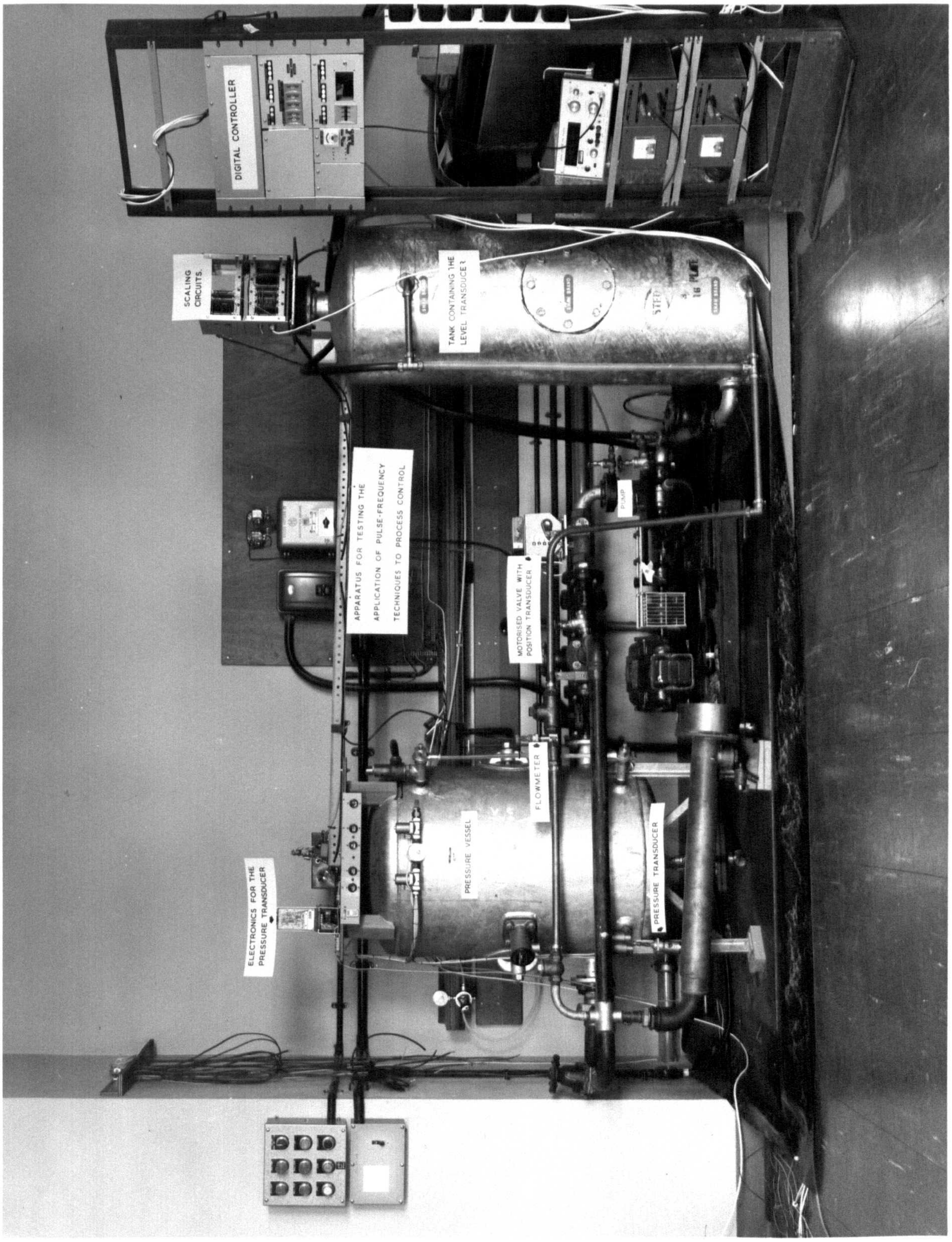


Photo A: Test rig for pulse frequency techniques.

CHAPTER 1

INTRODUCTION

This thesis describes an investigation into the use of a particular form of signal in automatic control. Control is the manipulation of energy in response to information, and usually entails the transmission and storage of information, computation and power amplification. The form of signal that is convenient for one of these activities may be unsuited to another of them. A given control system may therefore employ more than one form and consequently contain signal translation devices.

Signals are usually divided into the classes analogue and digital. Pulse frequency modulated signals strictly belong to the former, but show some of the characteristics of the latter. In particular, such signals can be handled by binary devices such as counters and logic gates. This intermediate position appeared to offer the possibility of combining the respective advantages of analogue and digital methods of control. Consequently it was decided to examine the basic control functions to see which of them could be performed by a combination of pulse frequency coding and logic circuits. A number of pulse-frequency techniques were known to exist, each having a specified field of application; what was hoped for was the creation of a more extensive control technology whereby pulse frequency signals could be used systematically.

Two analogue forms are widely used in process control; they are air pressure and electrical potential difference (or current). Transducers exist to convert a variety of variables into one or other of these, which are then suitable for transmission over moderate distances. Computation is readily performed in terms of these forms, as is power amplification. Although both pressure and voltage can be stored using appropriate 'capacities', such storage is rather ephemeral and other analogues such as position or resistance are also used.

Some disadvantages of analogue control have become apparent during recent years. Analogue controllers are not very accurate and are subject to long-term drift. Simple analogue communication over long distances (to a centralised control room) introduces considerable error into measurements. Analogue displays to human operators are difficult to read accurately or reliably if resolution is particularly fine, such as ½%. It is difficult to design or calibrate analogue controllers to perform such complex actions as interactive control, parameter adaptation, etc.

The falling price of general purpose digital computers has encouraged their use for control. Although they can only perform one task at a time, they are so fast that several control operations can be interleaved, sharing the same computational circuits. Information storage is easy, accuracy is high, display is numerical, and complex control algorithms can be performed. However the adaptation of a computer to a particular set of control tasks requires considerable skill. Modifications to control algorithms are difficult to perform 'on-line'. Placing many control functions, many loops, on a single device raises such problems of security that standby devices are often used. Digital encoders, such as digital transducers or analogue to digital converters, are rather expensive. The transmission of digital numbers over long distances requires complex terminal equipment.

In the foregoing paragraphs the merits and demerits of analogue and conventional digital control have been listed. A pulse frequency signal can be converted to and from analogue and (parallel) digital form fairly readily. This thesis describes the properties of such conversions and of both analogue and digital operations on pulse frequency modulated information. Chapter 2 treats the analogue operations and p.f.m. telemetry. Chapter 3 describes and evaluates the special serial digital techniques by which pulse-frequency and pulse-number signals can be processed. Chapter 4 is concerned with examples of control systems employing p.f.m. techniques.

During the period of this research, interest in p.f.m. techniques has been growing; as is shown in manufactured products, in the holding of an I.F.A.C. pulse symposium and in the publication of papers describing particular aspects. Despite this, the techniques are still treated on an ad hoc basis rather than as parts of a unified design method. This thesis stresses that unity.

Unfortunately there is some uncertainty about the meaning of the phrase 'pulse frequency modulation'. It is sometimes applied to a form of signal comprising bursts (or 'pulses') of oscillations of a variable frequency. That is not the subject of this thesis, throughout which 'pulse frequency modulation' is taken to mean the encoding of information linearly into the repetition rate of a train of events or pulses. Strictly, the form under consideration should be called 'unipolar integral pulse frequency modulation'.

CHAPTER 2

PULSE FREQUENCY MODULATION

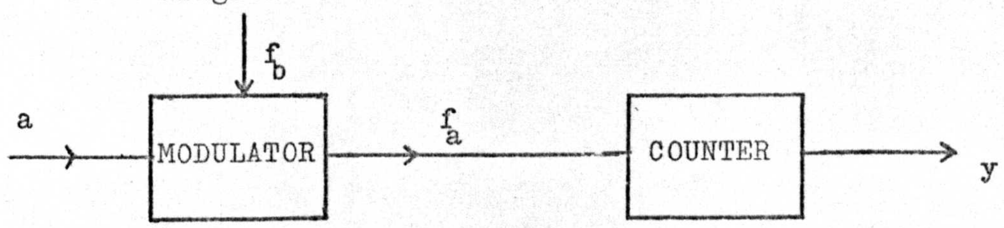
Contents:

2.1 Notation

This thesis covers many aspects of pulse-frequency modulation, and no single notation has been developed that is suitable for rigorous use throughout. The majority of symbols are redefined periodically, as appropriate to particular sections. Published work on pulse-frequency techniques has variously used notations borrowed from probability theory, signal theory, transform theory, descriptions of Markov processes and of incremental digital devices.

Martin[201] has been concerned to distinguish systematically between a quantity represented by a coded digital number and one represented by the repetition rate of a pulse train. For the quantity A he has used the forms $(A)_c$ and $(A)_r$ respectively. In a fairly narrow context his notation is satisfactory, but does not lend itself to situations where both variables and their Laplace transforms are to be considered.

In this thesis, variable quantities have generally been represented by lower case characters, frequencies by the letter f, Laplace transforms by corresponding upper case letters. For example a modulator followed by a pulse counter might be represented by the block diagram



The exact form of the signal put out by the modulator is complex and generally immaterial. It is some form of pulse train whose instantaneous rate is f_a or $f_a(t)$. So in the diagram, a (or a(t)) could be the instantaneous amplitude of a voltage; f_b and f_a (or $f_a(t)$) are frequencies; y (or y(t)) is a number. Representative equations in the time and transform domains would be:

$$f_a(t) = f_b(t) + K.a(t)$$

$$F_a(s) = F_b(s) + K.A(s)$$

K has units T^{-1}

and

$$y(t) = y(o) + \int_o^t f_a(x)dx + \text{rounding error}$$

$$Y(s) \doteq \frac{F_a(s)}{s} \qquad (2\text{-}1)$$

The notation just employed is somewhat inexact, for the instantaneous 'frequency' or 'rate' of a pulse-train is a conceptual quantity whose short term average is periodically observable. However it is useful to be able to handle quantities encoded as pulse-rates in terms of variables indicating the quantity rather than describing the signal waveform. The use of suffixed f has proved convenient.

The example just considered (the 'integration' of a pulse-frequency) raises another difficulty of notation or nomenclature. Taking sinusoidal signals, the time derivation of the phase is the instantaneous pulsatance, ω, and a cycle of the signal waveform is identified with a change in phase of 2π radians. With pulse-rate signals, the time derivative of a pulse count is the pulse-rate. A 'cycle' of the signal waveform is identified with a change in pulse-count (or pulse-number) of 1. Moreover the word 'phase' is commonly applied to pulse-frequency signals in a relative sense, one pulse-train can be 'in phase' or 'out of phase' with another.

Thus count or number in pulse-frequency work has much the same meaning as phase in sinusoidal signal work, excepting a factor of 2π in the units employed. Phase is a less ambiguous word than number or count, which have several other uses, and in this thesis 'phase' is used to describe the time integral of instantaneous frequency. It is presumed to be measured in counts or cycles, and not in radians, so that

$$f_a(t) = \frac{d\phi_a(t)}{dt} \quad , \quad \underline{\text{not}} \quad \frac{1}{2\pi} \cdot \frac{d\phi_a(t)}{dt} \tag{2-2}$$

and equation (2-1) might be rewritten

$$y(t) = \text{constant.} + \text{integer part} \left[\phi_a(t) \right]$$

$$\Phi_a(s) = \frac{F_a(s)}{s} \tag{2-3}$$

Other notations used throughout are:

P( ) for probability, p( ) for probability density

t for elapsed time, τ for pulse-period, T for counting period

f_b for bias frequencies, f_r for constant reference frequencies

N for number of stages, M for capacity of a counter

e for errors

σ_a^2 for variance of a

s for complex pulsatance (of Laplace transform)

2.2 Modulation

2.2.1 Integral pulse-frequency-modulation (IPFM) and related forms

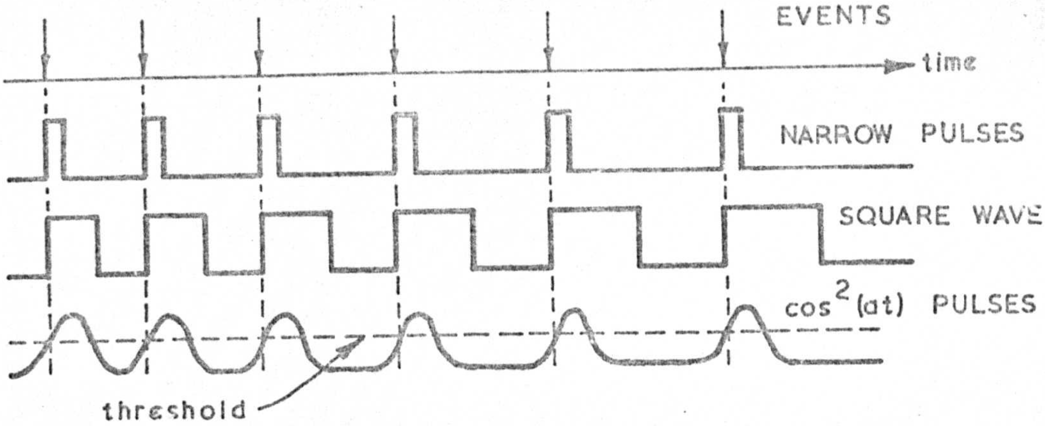
A point process is a succession of events whose times of occurrence are specified. The exact nature of the events is not determinate, except that the occurrence time associated with each event must be measurable. For signal analysis purposes, it is appropriate to replace the word 'event' by the word 'pulse'. In this context, then, a pulse is a signal of finite duration of such a form that it is possible to describe the time at which it occurs.

Pulse-frequency modulation is one technique for obtaining a point process from a continuous signal. It is governed by the rule that the time elapsed from one pulse to the next pulse is some linear function of the continuous signal. In its simplest form

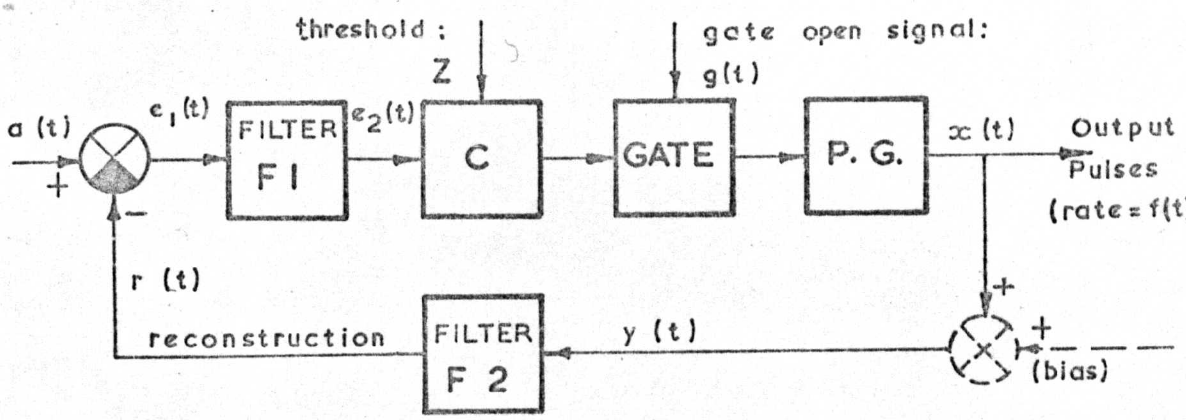
$$\text{Pulse rate} = \frac{1}{\text{Pulse period}} \propto \text{Continuous signal}\Bigg|_{\substack{\text{mean over}\\ \text{previous pulse}\\ \text{period}}}$$

In more complex forms, it is the <u>mean</u> pulse rate (over a number of pulses) that is related to the continuous signal.

The actual form of the pulses that comprise a PFM signal is open to variation. In a physical system they will be limited in energy and in bandwidth, but for purposes of analysis they are often conveniently treated as impulses, i.e. signals of infinite energy but negligible duration. The demodulation of pulse-modulated signals is an exercise in detection. Pulse shapes are therefore used which are particularly amenable to detection in the presence of noise or other causes of distortion. Sketched below are a set of possible signal waveforms expressing a set of event times:



In order to relate integral pulse-frequency-modulation to other forms of pulse modulation, the modulation process is represented by the following block diagram.



a (t) is continuous input

C is a comparator, P. G. a pulse generator

Fig 2A Modulation process:– block diagram

The input signal is assumed positive, as is the threshold of the amplitude detector.

With integral-pulse-frequency-modulation (IPFM)[202-207,210], the filter F2 is a simple gain element, while the filter F1 is an integrator, and the 'gate open' signal g(t) is always present. If the area of an output pulse is A (e.g. units of volt-seconds in an electrical system), and the gain of the filter F2 is K, then

$$\overline{a(t)} = K.A.\overline{f(t)} \tag{2-4}$$

where $\overline{a(t)}$ is the mean of the continuous input signal and $\overline{f(t)}$ is the mean of the output pulse rate. For waveforms in the converter, see Fig. 2B.

A synchronous form of IPFM is obtained if the gate is only opened very briefly at regular intervals. To achieve this, the gate open signal should be a series of pulses from a clock of repetition rate: f_r. The output pulses are now synchronised to the clock; that is, they are subject to quantisation in the time domain (see Fig. 2B). This results in substantial loss of information; the spacing of the output pulses is constrained to be an integer multiple of the clock period $1/f_r$. However it is still true that on average the output frequency is proportional to the input amplitude, as shown in equation (2-4). Synchronous IPFM is more convenient than normal IPFM when the pulse train generated is to be merged with, cancelled against or otherwise multiplexed with other similar pulse trains. It may also offer better prospects for detection at the far end of a noisy channel, because the distorted pulse-train need only be examined for the presence of pulses at the clock instants.

The name sigma pulse-frequency-modulation (ΣPFM) [204,207,208] is given to coding schemes in which (see Fig. 2A above) the feedback path filter F2 is a simple gain, and the forward path filter F1 has low-pass characteristics. Usually F1 is a first order lag circuit, with transfer function $1/1+sT$. The gate-open signal $g(t)$ may be continuous or pulsed[208], the latter yielding the synchronous form of ΣPFM.

A quite different type of modulation results if the feedback path filter F2 (of Fig. 2A) is an integrator while the forward path

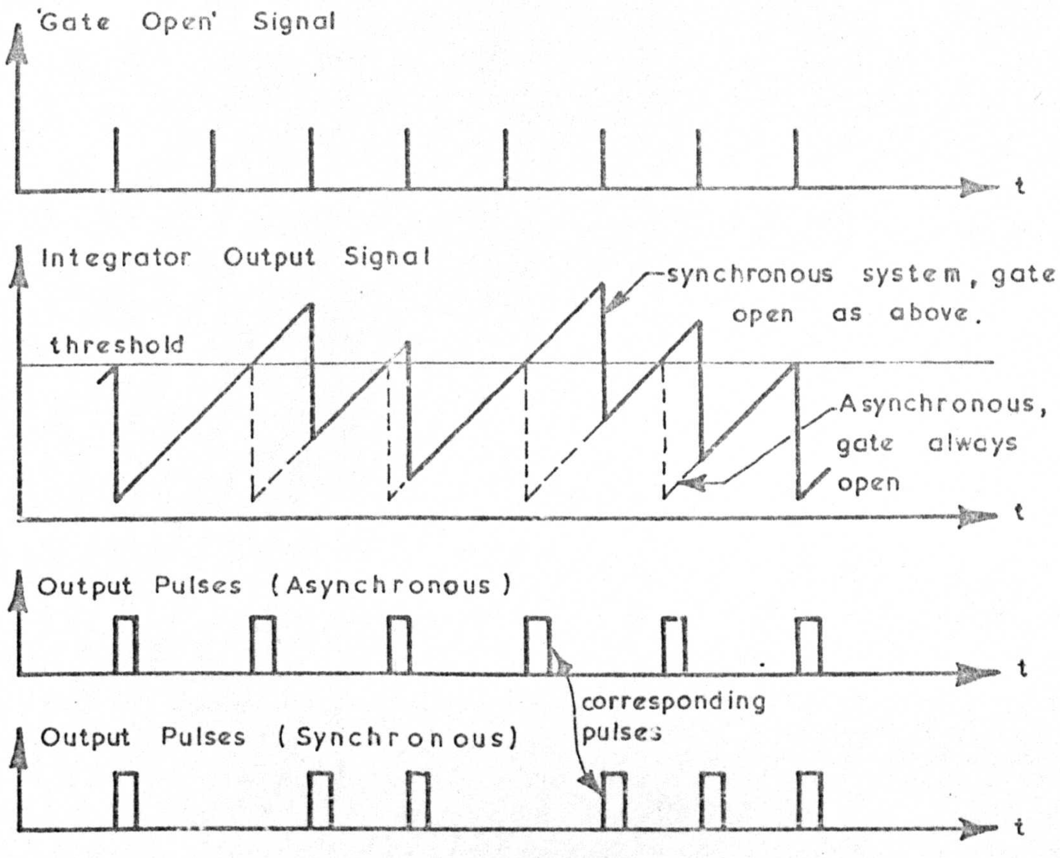


Fig. 2B Synchronous modulation (IPFM)

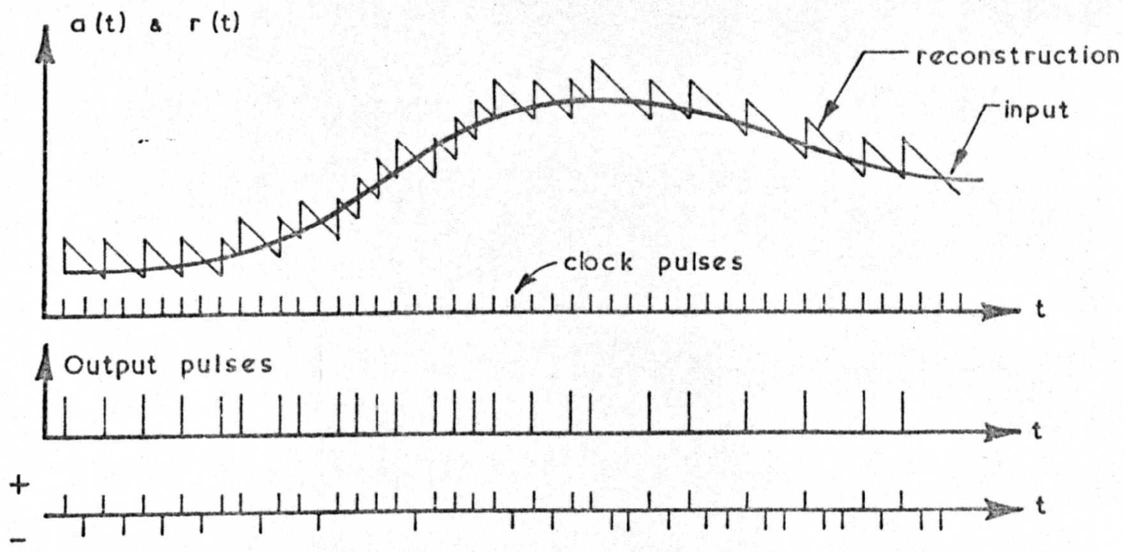


Fig. 2C Biassed synchronous incremental coding filter F1 is a simple gain. As in previous cases, the loop gain tends to minimise the difference $e_1(t)$ between the input signal $a(t)$ and the 'reconstruction' $r(t)$. In the case of IPFM and ΣPFM just considered, the output $x(t)$ is proportional to $r(t)$. In the present case, the output $x(t)$ is proportional to the time derivative of $r(t)$. The output pulses are thus related to changes in the input $a(t)$, and a constant input results in no output pulses. This way of encoding information is usually described as 'incremental'. If the input signal $a(t)$ is monotonically increasing, the gate open signal $g(t)$ is always present, and the bias (in the feedback loop) is zero, then an output pulse will be generated every time $a(t)$ increases by a certain amount. This amount, or step size, equals $K_i A$ where K_i is the gain of the integrator F2, A is the area of an output pulse.

A circuit that can only encode an ever increasing input has very limited use. A pulse tachometer, encoding shaft rotation into a pulse train, is one practical example. In general some means of handling decreasing as well as increasing inputs is required, and ternary incremental coding schemes are often employed.[211] Ternary coding can be avoided, or at least modified, by adopting the synchronous mode. In this mode, the gate of Fig. 2A is operated by a train of clock pulses, while a bias is introduced into the feedback path. This has the following effect: an output pulse occurring at a clock stroke represents an increment, while the absence of an output pulse at a clock stroke represents a decrement. In effect, the binary clock pulse-train is being subtracted from the binary output pulse-train to give a ternary coding. This is illustrated by the wave forms of Fig. 2C.

Incremental coding schemes are prone to drift, - for they depend upon an infinite memory. One lost pulse will result in a slight error (when the pulse train is decoded) that thereafter never decays. For this reason true incremental coding is only used where periodic error correction is possible, e.g. clock systems, machine tools, and where very reliable encoding devices are available. The transmission of speech does not require the handling of zero frequency components, and therefore a true incremental modulation is not suited to it. The use of a low pass filter[211,215] for F2 in Fig. 2A gives a quasi-incremental type of modulation wherein errors (due to loss of pulses) die away. This technique is called 'Delta' modulation (ΔPFM), and has been extensively developed. Abate[209] has reviewed progress up to 1967, while more recent work has been concentrated on methods to combat the effects of slope overload. As with other incremental codes, ΔPFM introduces errors due to amplitude quantisation and errors due to limitations on the gradient of an input signal that can be followed. In Fig. 3C for example, the 'reconstruction' signal r(t) is unable to follow the input a(t) when the latter is falling very rapidly. The reduction of these two types of error requires either a higher clock rate (and hence bandwidth) or conflicting constraints on the modulator gain. Various types of adaptive gain controls have been proposed[212-214] under such titles as 'high information' ΔPFM, 'companded' ΔPFM, to minimise the total coding errors for various types of input.

The various forms of modulation described above are subject to some variation in nomenclature .. for example the term 'Delta Sigma' ($\Delta\Sigma$PFM) is used to describe synchronous forms of both Integral (IPFM) and Sigma (ΣPFM) modulation. The distinction between one form and another is not always a sharp one. IPFM can be thought of as a limiting case of ΣPFM. Synchronous forms become indistinguishable from normal forms at very high clock rates. Slight variations of these main forms also exist. The interpulse period may be biassed upwards by a constant pulse-width[216]. The shaping of the outward pulse may be performed in the main modulation circuit rather than subsequently to it[217].

Where input signals are both positive and negative, either an input bias must be applied, or a ternary form of coding may be required. Considerable analysis has been applied to ternary (or bipolar) forms of IPFM and ΣPFM; however these forms require rather elaborate circuits and generate pulses requiring fairly complex detection circuits.

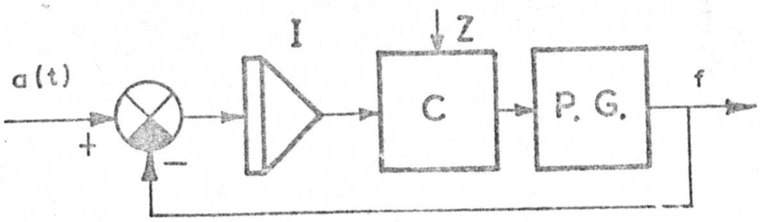
The principal features of the unipolar forms of pulse-modulation are shown in Table 2.1. These forms all conform to Fig. 2A, and therefore exclude certain other point-process generating methods such as pulse-position modulation. To reiterate: this thesis is concerned with the simplest of the modulations, viz. unipolar integral pulse-frequency modulation, IPFM.

Table 2.1 (Refer to Fig. 2A)

Name of modulation	Abbreviation	Fig. 2A				Comments
		F1	F2	Bias	g(t)	
Integral	IPFM	$\frac{K_i}{s}$	K	No	Contin.	Simple circuits.
Synch. Integral	Synch.IPFM ($\Delta\Sigma$PFM)	$\frac{K_i}{s}$	K	No	Clock	For multiplexing or computation
Sigma	ΣPFM	$\frac{K_i}{1+sT}$	K	No	Contin. or clock	Can be matched to particular class of input.
Incremental		K	$\frac{K_i}{s}$	No	Contin.	Ternary. Error prone.
Synchronous Incremental		K	$\frac{K_i}{s}$	Yes	Clock	Binary. Error prone.
Delta	ΔPFM	K	$\frac{K_i}{1+sT}$	No	Contin. or clock	Voice or video telecomms.Slope limited. Adaptive forms.

2.2.2 Modulation Circuits

As discussed in the last section, the modulation circuit for IPFM may be represented by the following diagram:



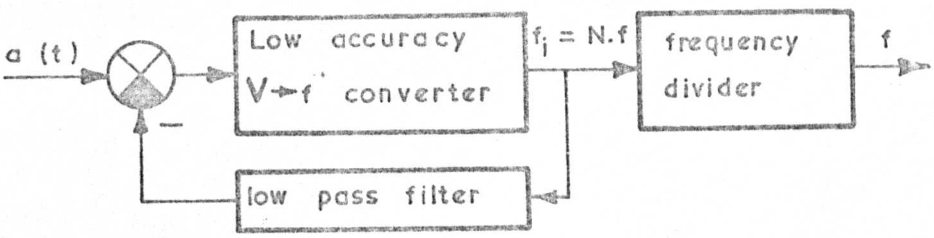
I = Integrator

C = Comparator

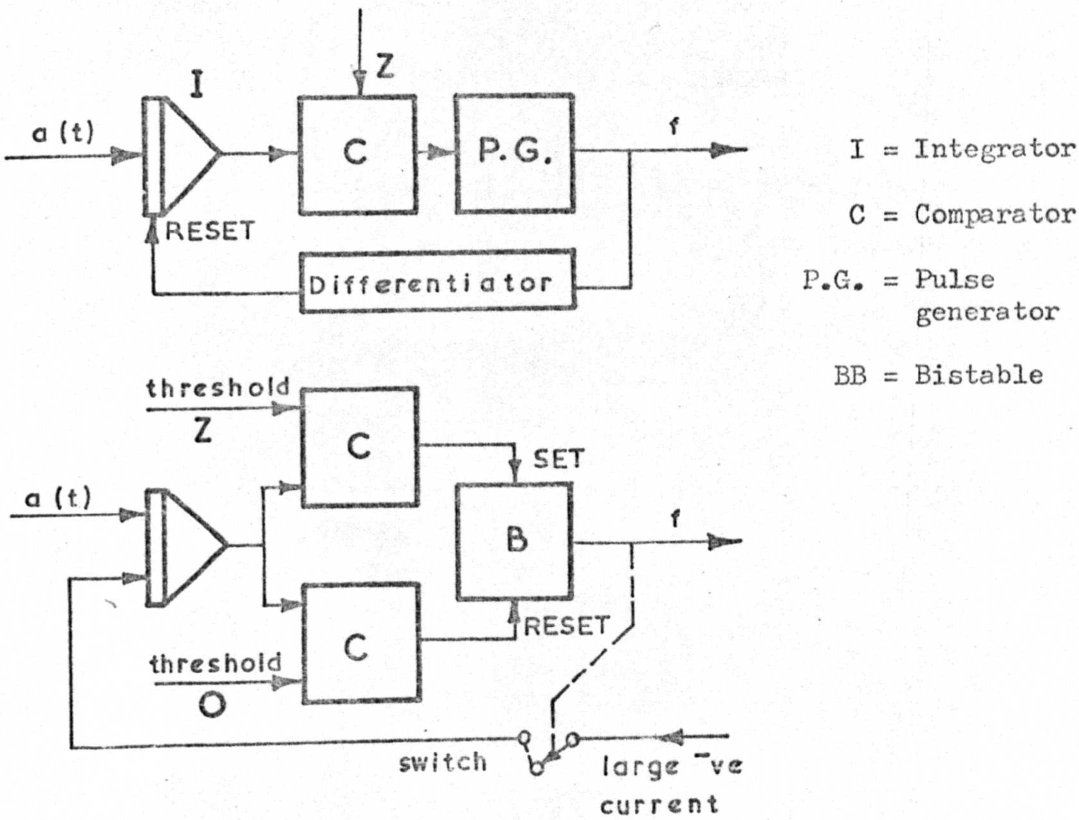
P.G. = Pulse Generator

Because of the feedback, the accuracy of modulation depends primarily on the accuracy with which output pulses are generated and upon drift in the integrator. Integrator gain can vary slowly without affecting accuracy. The output pulses should have closely defined areas (i.e. length x height product for rectangular pulses); The integrator drift should be equivalent to an input offset of less than the desired resolution of a(t). Using standard linear integrated circuits, it is not difficult to achieve modulation accuracies of 1%. Accuracies of 0.1% require the use of precision components and some restriction on ambient temperature range. Controlling integrator drift is usually fairly easy, defining pulse areas is less so. With synchronous IPFM, a fixed frequency 'clock' can be used to define pulse-length more reliably than when using a monostable. Pulse-height definition by the use of single or cascaded zener diodes is most easily achieved for 5 volt pulses, because at such levels zener diodes have low temperature coefficients.

A variant of an integral pulse-frequency modulator has been proposed[218] wherein an open-loop voltage-to-frequency converter replaces the integrator-comparator arrangement. Accuracies of 0.1% are claimed. Because of poor dynamic response, it would be advisable to generate a high internal frequency f_i and divide down to obtain an output frequency in the desired range. Such a modification of the original proposal is shown in the sketch below. The cut-off frequency f_c of the low-pass filter should satisfy $f_i > f_c \gg f_a$



The block diagram shown at the beginning of this section may be approximated by arrangements where the integrator output is very rapidly reset to zero each time an output pulse is generated. This resetting is a replacement for the normal feedback path. Two possible arrangements to reset the integrator are shown below.



In both cases the correct output interpulse period p, that satisfies

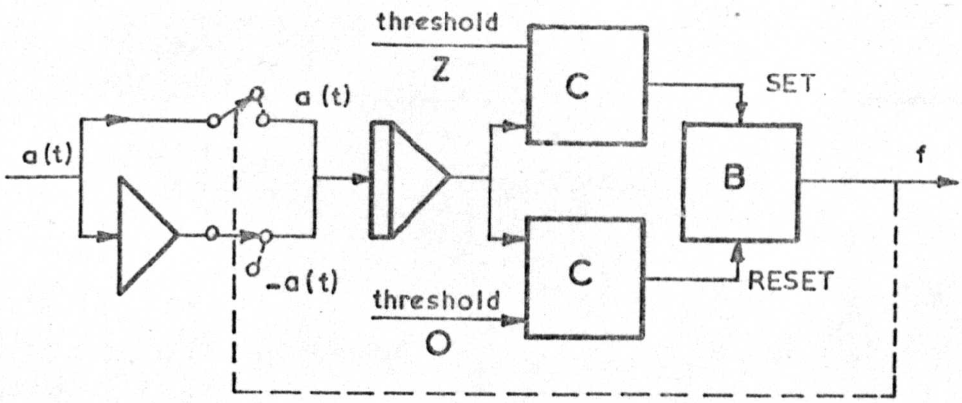
$$Z = K_i \int_{t_i}^{t_{i+1}} a(t)dt \qquad \text{where } t_{i+1} - t_i = p_i(t) \qquad (2\text{-}5)$$

is incorrectly increased by the resetting time T_r. So for acceptably small errors

$$T_r < (p)_{min} \times \text{resolution required.}$$

The resetting of an integrator requires the discharging of a capacitor, so that a lower limit is placed on T_r by the maximum discharge current available from the integrator output, or through a fast switch such as a field effect transistor.

A further variant employs a dual ramp technique with two comparators. This gives a square wave output. An accurate inverter, accurate integrator and high speed switches are required. The circuit is shown below.



In each of the last three arrangements, the need to generate accurate pulses is obviated, but at the expense of increasing the accuracy in the integrator time constant.

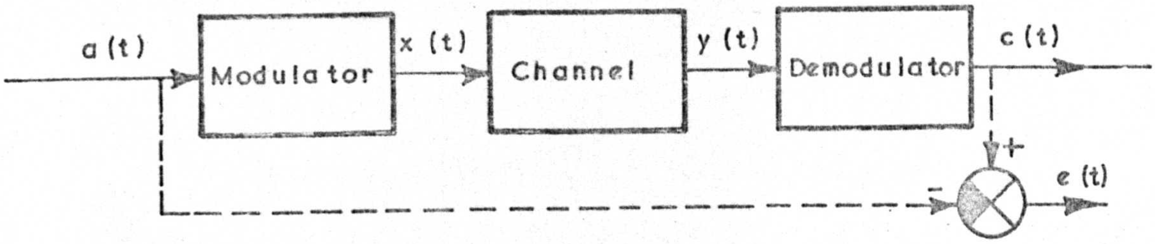
The circuits described so far are based on high gain amplifiers used as nearly ideal integrators. Much has been written about much simpler voltage-to-frequency convertors employing unijunction transistors[219] or small numbers of bipolar transistors. The circuits generally used as (non-pulse) frequency modulators are prone to drift and are thus unsuitable for transmitting zero-frequency signal components. They are moreover usually only linear over a limited range wherein the instantaneous frequency is within a few per cent of the carrier frequency.

Circuits based upon bistable transistor pairs have been very popular and refinements have been developed to increase the basic accuracy and range. The state of the art is represented by Korytowski[220] (accuracy), Tesic[221] (range), Smith & Sedra[222] (range and accuracy), Marconero & Pallottino[223] (6 decade range). Further development of bistable circuits would not seem appropriate, now that integrated circuit operational amplifiers are cheap and effective. An integrated circuit voltage-to-frequency converter is not an immediate possibility due to difficulties in forming capacitors of sufficient stability and size. Hybrid circuits would seem more promising.

Non-linear modulators have also been invented: a useful one by Bamford[224] generates a pulse-frequency that varies with the square root of the input signal amplitude.

2.2.3 Spectrum, transmission properties

Using PFM, a time varying signal (= transducer output = modulation signal = baseband signal) a(t) is converted to a transmission signal x(t) that is passed into a communication channel. At the far end of the channel, the received signal y(t) is converted into an output c(t) which is generally similar to the input, but differs from the input by some error e(t).



The transmitted signal is a sequence of well-shaped pulses: the received signal pulses are less well shaped. The design of the modulator and demodulator should minimise the error for some statistically specified channel characteristic. Considerations of expense usually rule out 'ideal' modulation and demodulation circuits, and it is then necessary to identify designs that are 'good enough'.

The observed error e(t) will have components due to imperfections in the modulator, in the channel and in the demodulator. Channel imperfections will generally be compounded by those in the demodulator.

An appreciation of the general communication properties of PFM may be obtained by examining the power spectrum of the transmitted signal. A train x(t) of evenly spaced pulses of rate f_b, each of height h and width w, may be expressed by the Fourier series:

$$x(t) = hwf_b \left\{ 1 + 2\sum_{n=1}^{\infty} \frac{\mathrm{sinc}(n\pi wf_b).}{n\pi wf_b} \cos(2\pi nf_b t) \right\} \qquad (2\text{-}6)$$

The corresponding spectral powers are:

$$\begin{aligned} &\text{at frequency} = 0 &&: H^2 \\ &\text{at fundamental frequency} = f_b &&: H^2.2\mathrm{sinc}^2(\pi K) \qquad (2\text{-}7)\\ &\text{at harmonic frequencies} = nf_b &&: H^2.2\mathrm{sinc}^2(n\pi K) \end{aligned}$$

where $K = wf_b$ and $H = hK$

The effect of pulse width on power distribution may be observed from the following table. In the table very narrow (K = .01), narrow (K = .1) and wide (K = .5) pulses are considered.

Spectral power when pulse amplitude (h) = 1, and $K = \dfrac{\text{pulse width}}{\text{pulse period}}$.

K	P(0)	Power Ratio $P(nf_b)/P(0)$						
		n = 1	2	3	4	5	6	7
0.01	0.0001	2.00	2.00	1.99	1.99	1.98	1.98	1.97
0.1	0.01	1.90	1.74	1.46	1.21	0.81	0.51	0.27
0.5	0.25	0.79	0	0.09	0	0.03	0	0.02

For very narrow pulses, the overall signal power is low, but is spread across a wide frequency band, and fundamental frequency power is double that at zero frequency.

For wide pulses, the overall power is naturally high. This power is concentrated in the zero-frequency and fundamental-frequency components, rather more power in the former than the latter.

When the pulse frequency is modulated by an input signal, the power spectrum just described becomes modified. The 'lines' at $f = 0$, f_b, $2f_b$, $3f_b$, etc. broaden into bands centered upon these frequencies.

Fig. 2D shows the Fourier coefficients of the transmitted signal x(t) from an integral pulse-frequency modulator whose modulating signal is $a(t) = A \cos(2\pi f_a t)$. There is a frequency bias, so that instantaneous frequency: $f = f_b(1 + A \cos(2\pi f_a t))$. In Fig. 2D $f_b = 250$ Hz, $f_a = 1$ Hz and $A = 0.16$ and the pulses are very narrow. The spectrum shown is a discrete one, calculated for frequency increments of 1 Hz corresponding to the spacings of the line spectrum of a(t) and hence x(t). The spectrum is arbitrarily scaled to give a zero frequency component of 100.

From Fig. 2D the following features may be noted:

(i) The spectrum of a(t) appears in that of x(t), so that a low-pass filter would enable a(t) to be recovered from x(t). However, most of the low-frequency power is concentrated at zero-frequency and is due to the bias of 1 added to the input. This presence of the modulating signal as a component of the transmitted signal is a distinctive feature of PFM and not often encountered elsewhere in modulation theory. The feature is of little use when signals are multiplexed into a single channel, or when the channel has poor low-frequency characteristics. However for undistorted transmission signals, very simple (e.g. R-C) circuits can be used for demodulation.

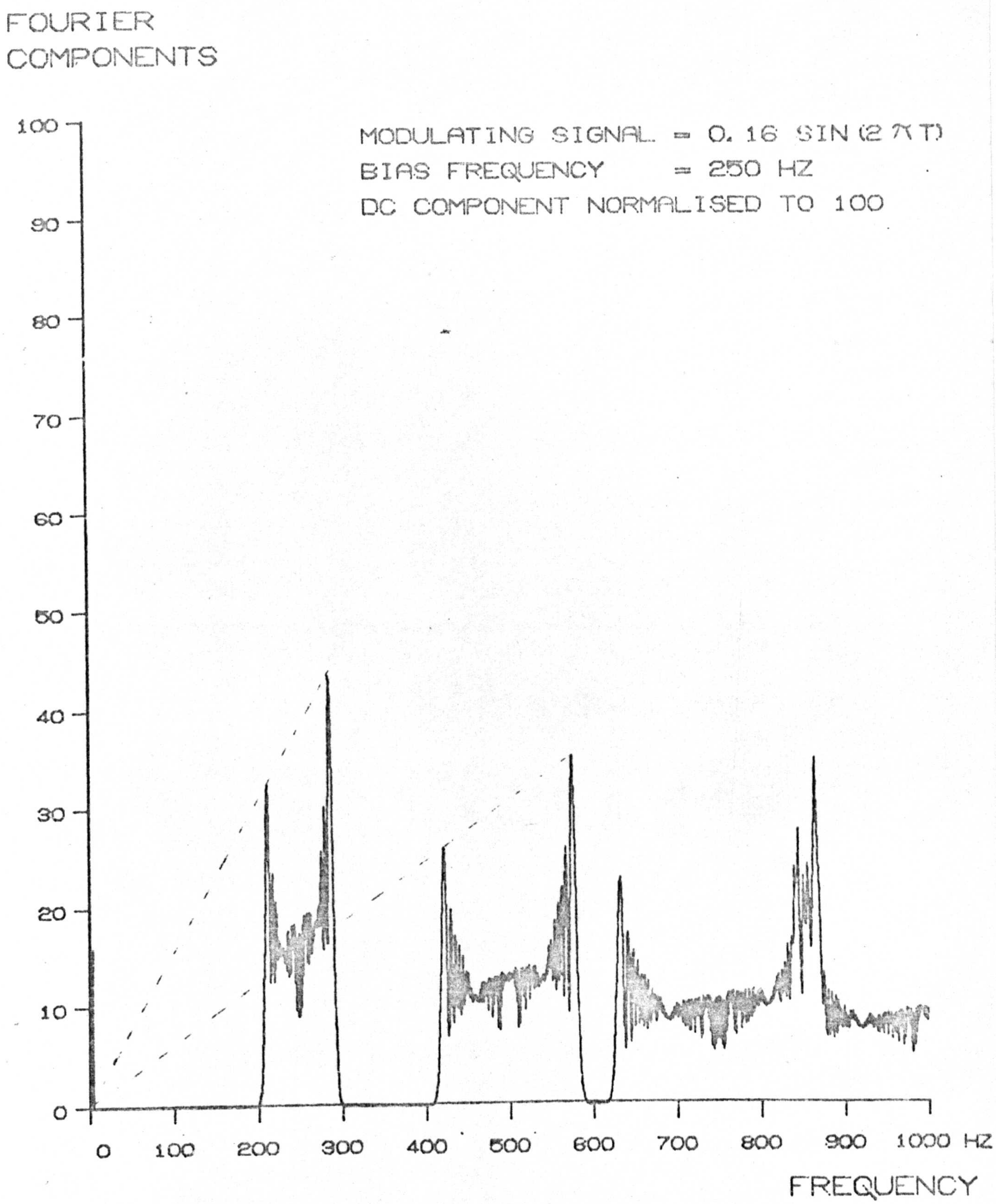


FIG. 2D PFM SPECTRUM, SINUSOIDAL MODULATION

(ii) The line at frequency f_b (= 250 Hz) is widened to a band of width about 90 Hz. This band is very similar to the entirety of a frequency modulated signal, although in PFM it forms only a part. The power in this band is generally somewhat greater than that at low frequencies - as discussed earlier. The modulation index (in its FM sense), the ratio of maximum instantaneous frequency deviation to modulating frequency, equalling 0.16 x 250 Hz/1 Hz = 40, is quite high. Consequently[227] the sidebands about 250 Hz fall in amplitude quite rapidly outside the range of instantaneous frequency: 210-290 Hz. Because the ratio of carrier to modulating frequency is very high, the shape of the spectrum about f_b is approximately that of the amplitude probability density of $a(t) = A\cos(2\pi f_a t)$: cusp-shaped tending to infinity at the cusp edges. The fundamental band spectrum of a PFM signal differs from the spectrum of an FM signal in being skew. The PFM band can (approximately) be obtained by weighting each component of the corresponding FM spectrum by the factor: component frequency/f_b. Thus the upper sidebands are accentuated. With square wave frequency modulation, an intermediate form between FM and PFM, this skewing would not occur.

(iii) Bands of width proportional to centre frequency are present about each multiple of f_b, i.e. at 500, 750, 1000 Hz etc. Because bandwidth increases with frequency, whereas band spacing does not, the high frequency bands overlap.

The most significant part of the PFM spectrum may be considered to be the band about frequency f_b. The low-frequency band is an auxiliary one which is easily corrupted (being a form of amplitude modulation). The harmonic bands serve to sharpen the edges of the pulses. For reasons of crosstalk reduction, or inconstant channel group velocity, it is often necessary to filter out these higher frequency bands.

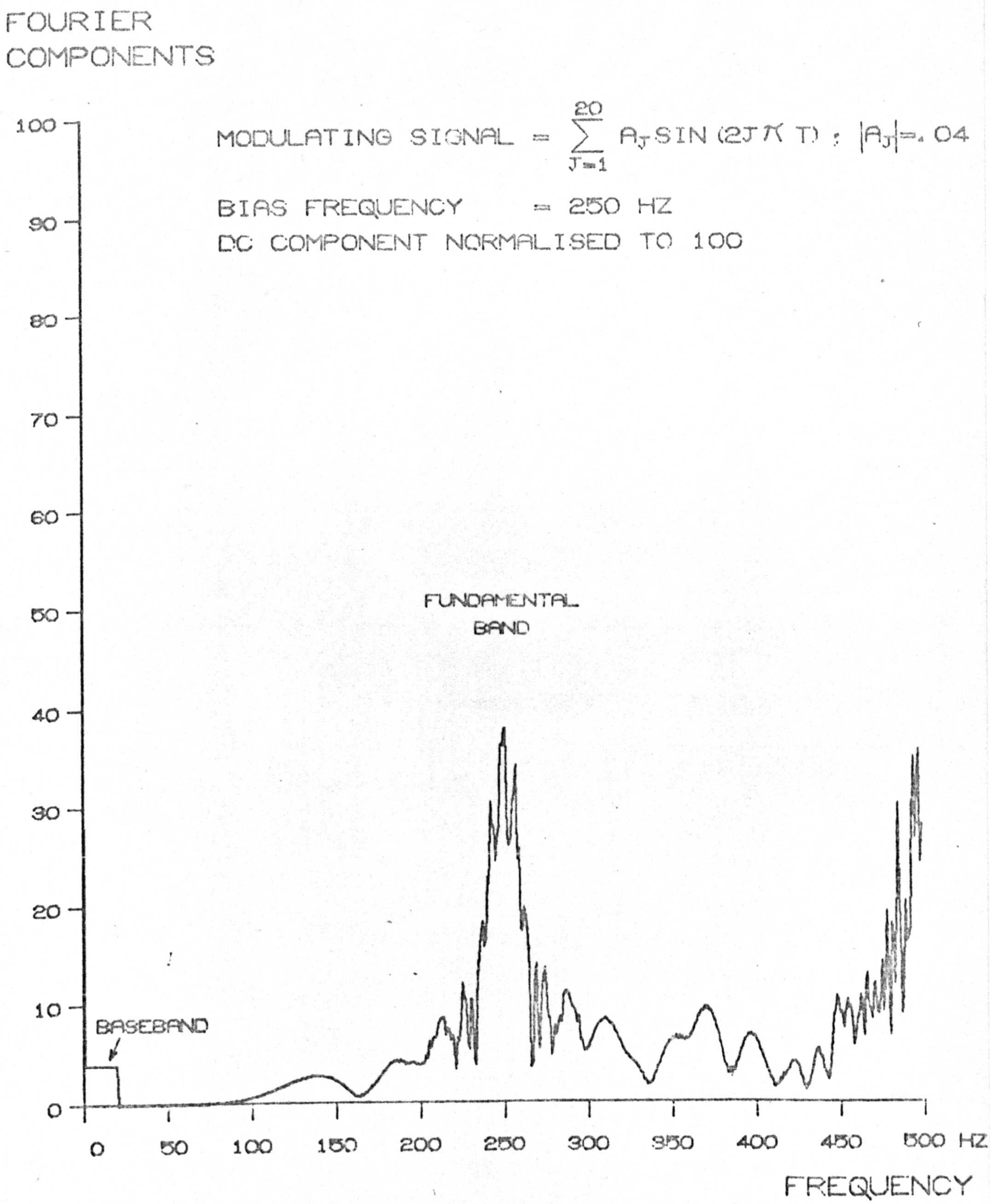


FIG. 2E PFM SPECTRUM, COMPLEX INPUT

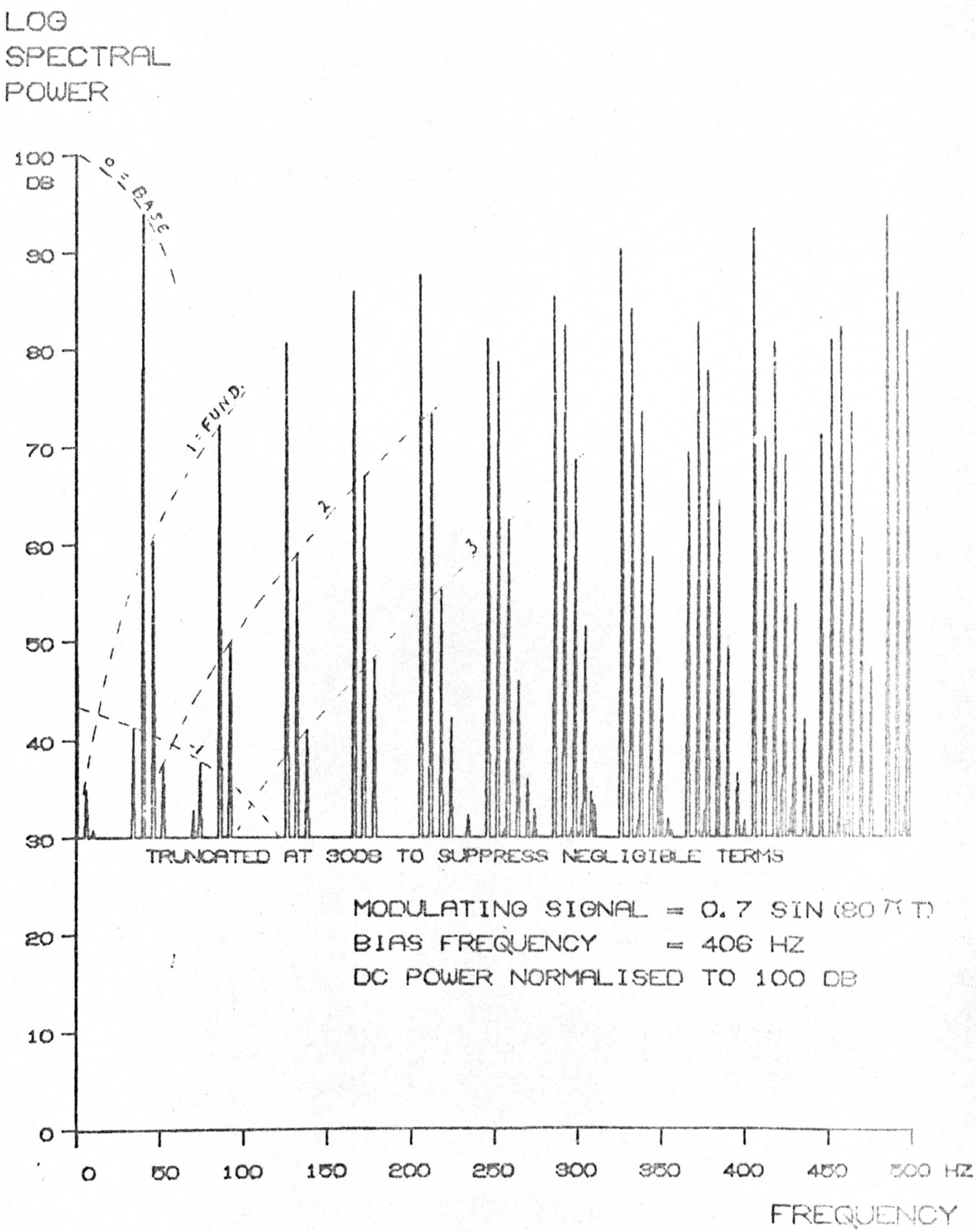


FIG. 2F PFM SPECTRUM, HIGH MODULATION INDEX

Overlapping of higher frequency bands has already been mentioned. It is not significant. Overlapping of the baseband and fundamental band is more critical. It removes the possibility of demodulation by low-pass filtering, and may reflect loss of information. Shannon's sampling rule predicts loss of information when the instantaneous sampling rate falls below twice the frequency of the highest-frequency spectral component of the sampled signal. Blanchard[225] examined the application of Shannon's rule to PFM and showed it to apply.

Fig. 2E shows the spectrum of the transmitted signal when the modulating signal has a band rather than a line spectrum, and normal rather than cusp-shaped amplitude probability distribution. Fig. 2F shows a log-power spectrum for a very large modulation index where the fundamental band begins to overlap the baseband.

The transmitted signal from a pulse frequency modulator has a very wide bandwidth extending from zero frequency. It is not suitable for radio transmission unless all bands except that about f_b are filtered out, in which case a signal akin to a frequency modulation signal is obtained. Removal of all bands except the baseband and the fundamental band (about f_b) gives a train of rounded pulses. Each harmonic band that is now replaced will tend to steepen and sharpen the pulses. In the presence of noise, sharp fronted pulses are defined in time with less error than pulses with slow rise times.

The minimum bandwidth required to transmit the baseband and fundamental band will be at least four times that of the modulating signal (= baseband). In most applications bandwidths of more than 50 times the baseband width are used. PFM is generally somewhat wasteful of bandwidth. It is most suitable for use where the signal to be transmitted has a bandwidth that is only a small fraction of the channel capacity. This often arises in process control.

2.3 Demodulation

2.3.1 Direct demodulation

The simplest demodulation circuits are those that act as low pass filters, separating out the baseband portion of the received signal. The cut-off frequency of such filters should lie between the top of the baseband (or modulating signal) and the bottom of the fundamental band of the received signal. The attenuation and phase shift should be negligible over the baseband; the attenuation should be high at the lowest instantaneous pulse frequency to minimise ripple. These conditions are difficult to meet unless the bias or carrier frequency f_b is many times the highest modulating frequency f_a. Using a first order RC filter, the specification that demodulation error due to lags or ripple should not exceed 5% requires $f_b > 1300\ f_a$. Using higher order filters the conditions are much less severe.

A circuit that uses more of the incoming signal's information than a low pass filter is the diode pump[228]. The input pulses are amplified and limited, and each reformed pulse 'pumps' a fixed charge Q into a capacitor. The capacitor discharges continuously through a resistor R. At equilibrium the potential across the capacitor is fQR, while the fractional ripple DV/V is 1/CRf. As with low pass filtering there is a conflict between immediate response to changes in received frequency f (small value for C). Also as with filtering use of a higher order diode pump[229] reduces this conflict.

Where a PFM signal is filtered (deliberately or in the process of transmission) to remove all except the fundamental spectral band about f_b, then an amplitude distorted FM signal results. The amplitude distortion can be largely removed if the receiver is provided with an automatic gain control of appropriate speed. Numerous circuits exist for demodulating FM signals, discriminators being most commonly used. These circuits depend upon one or more non-linear elements and are not very accurate, nor capable of handling signals with a high modulation index.

The properties of FM have been extensively studied. The property of greatest interest here is the reduction of the distortion in the demodulated signal, compared with direct or AM transmission over the same fading or noisy channel. Despite the greater noise content of the received signal of an FM system over an AM system, after demodulation the noise content is reduced by a factor up to $3 \times (\Delta f/B)^2$. Δf is the maximum instantaneous frequency deviation, while B is the bandwidth of the conveyed information. The reduction is with respect to AM using the same transmission power. Frequency discriminators fail completely, however, if the received signal-to-noise ratio after filtering falls below about 10[227,Chap.6]. Phase locked loops can be used where signal-to-noise ratio is as low as 3[230], provided the modulation index is very low. Phase locked loops can be made more stable and linear than discriminators, achieving accuracies of up to 0.1%.

2.3.2 Noise and error characteristics of PFM demodulation

Noise acquired by a signal being transmitted in an industrial environment may be divided into three categories

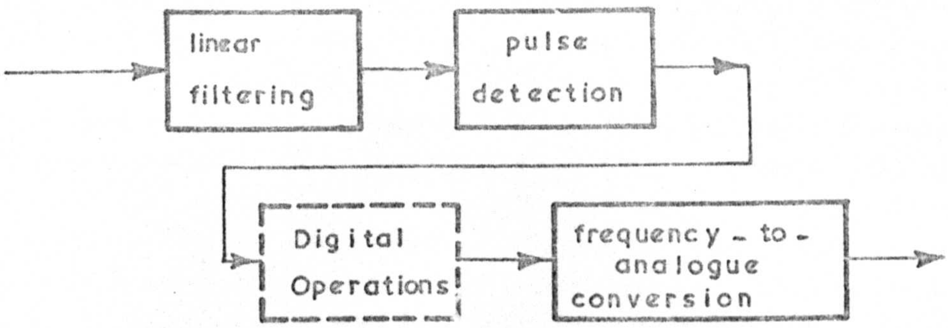
(i) d.c. bias or drift, due partly to earth leakage current and partly to uncompensated attenuation caused by line resistance;

(ii) substantial induced voltages at mains frequency and at the low harmonics of mains frequency, due to inductive and resistive coupling with high power equipment:

(iii) wideband noise due to crosstalk and capacitative or resistive coupling generally; the amplitude distribution of this noise is not always Gaussian, as large narrow spikes are not uncommon.

Directly transmitted information is particularly susceptible to corruption by the first two categories, FM by the last category only, PFM by all three but to a lesser extent. The multiple coding of information in PFM allows the detection circuits to be adapted to the noise characteristics. A full analysis of the effects of non-Gaussian noise with a very coloured spectrum is beyond the scope of this thesis.

PFM (as opposed to FM) transmission is generally associated with digital processing of the modulating information. For this reason interest centres upon demodulation that includes pulse-shape restoration. The final signal form may be analogue or digital.



Pulse detection is subject to two kinds of error, dependent on the degree of corruption of the received signal. Low noise levels will cause pulse-position errors. Large noise levels will also result in missed pulses or false pulses. It is the latter that are of prime interest, as is indicated by the ensuing analysis.

The pulse-position errors experienced with no loss of pulses are unlikely to exceed $\frac{1}{4}$ pulse period: so that the detected phase (in counts) may be written

$$\phi = f_b.t + f_b.\int^t a(v)\,dv + n(t)$$

$$\text{where } |n(t)| < \tfrac{1}{4}$$

The phase error n(t) has a spectrum extending fairly uniformly from d.c. to $f_b/2$. However, only spectral components with frequencies below about $f_b/20$ are significant, as higher frequency components will be removed during frequency-to-number (or frequency-to-voltage) conversion.

Assuming as a worst case that 10% of the peak amplitude of n(t) is due to its low frequency components, and that those are represented by a single sinusoid of frequency $f_n \ll f_c$. Then:

$$\phi = f_b.t + f_b\int^t a(v)dv + \frac{1}{40}\cos 2\pi f_n t + \text{high freq. terms}$$

The instantaneous frequency is:

$$f = \frac{d\phi}{dt} = f_b(1 + a(t)) - \frac{\pi}{20} \cdot f_n \sin 2\pi f_n t + \text{high freq. terms}$$

$$= f_b(1 + a(t) - \frac{\pi f_n}{20 f_b} \sin 2\pi f_n t) + \text{high freq. terms}$$

Information lies in the low frequency modulating signal a(t) whose maximum amplitude will be about $\frac{1}{2}$. Distortion lies in the low frequency term $\frac{\pi f_n}{20 f_b} \sin 2\pi f_n t$ whose maximum amplitude will be less than 0.01 when $f_n < f_b/20$.

So provided the modulation index is high (and values in excess of 100 are commonly used), pulse <u>position</u> modulation due to noise will not introduce significant errors in the demodulated signal. The important errors are those due to missed or false pulses, and depend on error rates in detection.

2.3.3 Detection techniques

A train of pulses has a wide but coherent spectrum: the coherence helps distinguish the pulse from interference containing a similar range of frequencies. This coherence or phase relationship is liable to distortion by imperfections in the transmission channel. For time-varying channels the distortion cannot be reliably predicted. For stable channels, phase distortion can be predicted from measurements, but these are often laborious. In the majority of practical cases, high frequency components that are subject both to attenuation and to complex phase distortion are removed by filtering, as they can only with considerable difficulty be used to aid detection. Only with very high capacity synchronous channels is the calculation of compensating filters (to remove phase distortion) thought worthwhile.

The one channel variation that can be compensated fairly simply is change in low-frequency gain. A detector may be preceded by an automatic gain control circuit with an appropriately long time constant.

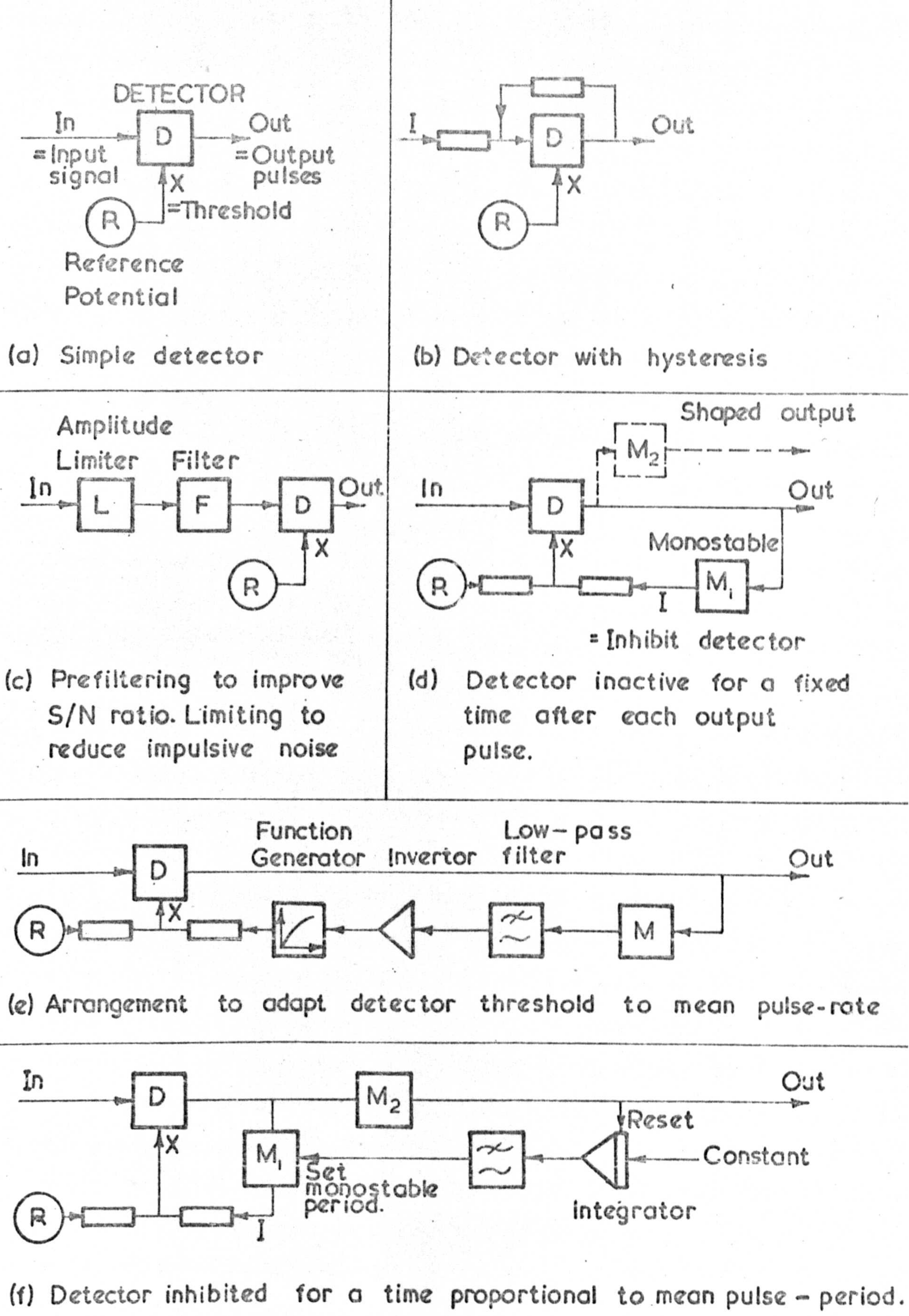
The principal techniques available for pulse-detection are listed diagrammatically in Fig. 2G. They are shown in order of increasing complexity, and of course the sequence may be extended by combinations of two or more techniques. The techniques make use of a priori knowledge of the pulses, or of the forms of distortion. This knowledge may comprise expectation of pulse-shape, of the size and spectrum of the pulse-modulating signal and of the statistics of additive noise. A balance has to be struck between optimum detection that uses all this a priori knowledge, and cheap detection that employs simple circuits.

Fig. 2G (b) shows the simplest way of using knowledge of pulse amplitude and rise-time. For a given noise, slow rise-times justify wider hysteresis bands.

Prefiltering, as in Fig. 2G (c), is only really effective where noise statistics are reliable. Much has been written about the filtering of signals from Gaussian noise, be it white or coloured. For event detection in white Gaussian noise the 'matched filter' (whose impulse response is the mirror image of the event pulse) is optimum. Unfortunately the larger components of electrical noise in industrial environments do not generally have a Gaussian amplitude distribution. Moreover 'matched' filters are generally difficult to construct, and second or third order approximations have to be used.

An experiment was carried out to examine the sensitivity of detection errors to filter design. Rectangular pulses in Gaussian wideband noise were generated, filtered and detected; the whole process being simulated in a digital computer. The matched filter 'A' (which is here a running averager) and a suitable second order filter 'B' were compared. Properties of the two filters are illustrated by Fig. 2H, from which the matched filter appears considerably superior. Measured performance figures are:

Noise at filter input in arbitrary units		1	2	4	8
Probability of a false detection	A B	.007 .010	.035 .050	.09 .12	.13 .19
Normalised standard deviation of entries above		.6	.3	.2	.2



(a) Simple detector

(b) Detector with hysteresis

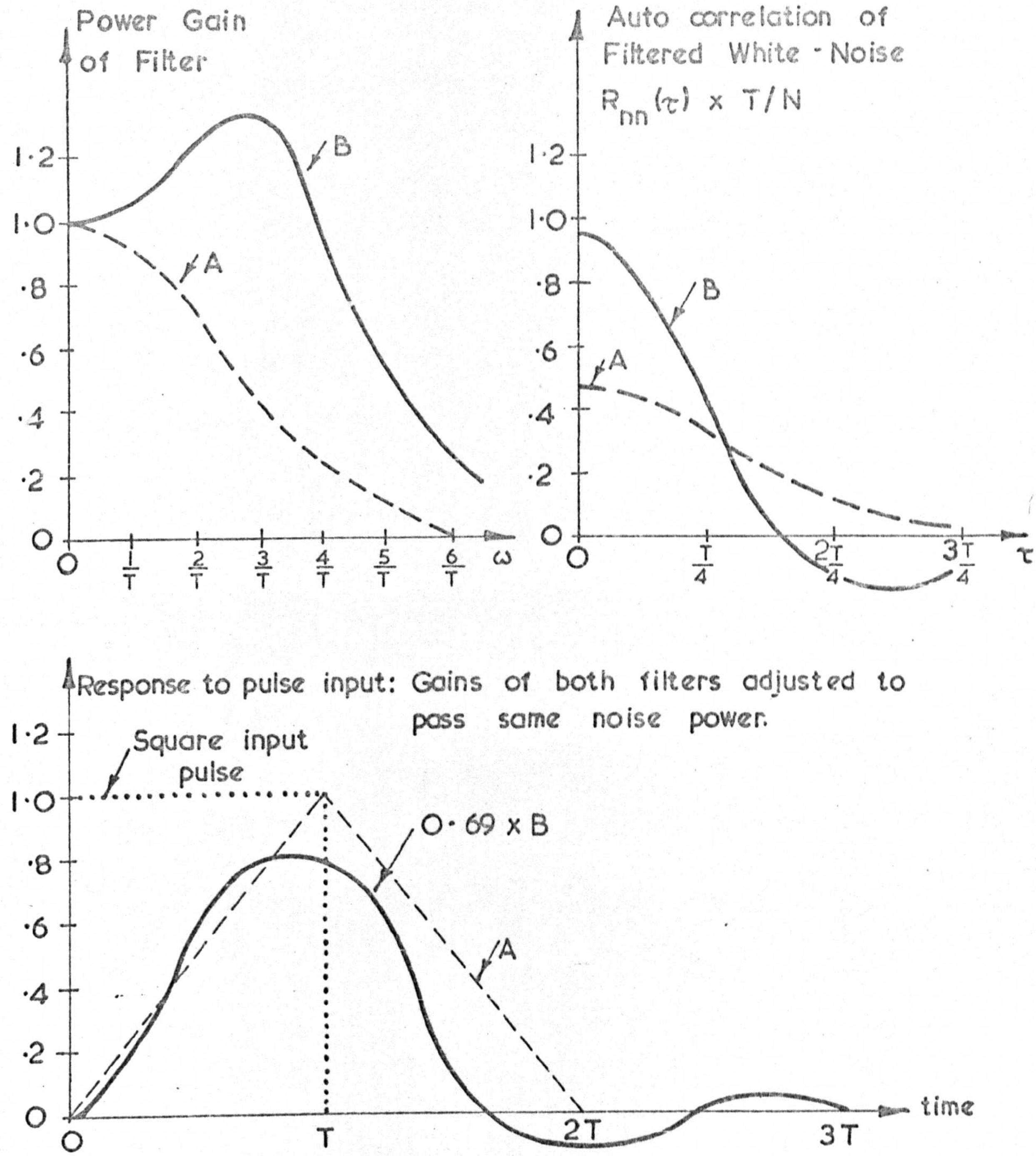
(c) Prefiltering to improve S/N ratio. Limiting to reduce impulsive noise

(d) Detector inactive for a fixed time after each output pulse.

(e) Arrangement to adapt detector threshold to mean pulse-rate

(f) Detector inhibited for a time proportional to mean pulse-period.

FIG 2G PULSE DETECTION CIRCUITS.



A - Matched filter, T.F. is $\frac{1-e^{-ST}}{ST}$

B - 2nd Order filter, T.F. is $\frac{1}{1+2K\,S/\omega_1+(S/\omega_1)^2}$ $\quad K = 0{\cdot}5$, $\omega_1 = 4/T$

FIG 2H COMPARISON OF NOISE AND SIGNAL RESPONSES OF TWO PRE-DETECTION FILTERS.

Thus the use of the simple filter 'B' has resulted in about 40% greater error probability than would apply with the matched filter 'A'. Such a percentage cannot be considered very significant in this context.

Non-linear filters may be optimum for non-Gaussian noise; they are extremely complex to design. Amplitude limiting circuits are helpful in the presence of impulsive noise spikes, and also serve to protect subsequent circuits from electrical damage.

Where channels are subject to slow drift or mains-frequency pickup, bandpass filtering may improve the signal-to-noise ratio. However, the presence of the appropriate low-frequency signal components allows relatively simple detectors to be used, and it is sometimes worthwhile restoring these components after filtering, by feedback from the detector output. Fig. 2G (e) shows one method of achieving this restoration by adjustment of threshold level. In fact variation of this level is helpful even when the pulse-train has not lost its low-frequency component. Using the recent past pulse-rate as an indication of current pulse rate, the threshold can be varied to maintain the expectation of false pulses and missed pulses approximately equal.

For example, with rectangular pulses of width T and unit height, in Gaussian white noise, using a matched filter and a detector inhibited for time 1.8 x T after each detection, the following table was derived:

Pulse period	2T	3T	4T	5T
Measured optimum* threshold	0.33	0.53	0.62	0.64
Threshold depression to compensate for loss of d.c. component	0.50	0.33	0.25	0.20

* i.e. + and - errors equalised

Even when the threshold is optimum, errors rise as pulse-rates fall. The detector of Fig. 2G (d) makes use of knowledge of maximum pulse rate, by turning the detector off for a fixed time after each detection. By experiment the optimum time was found to be about 90% of the minimum expected pulse period. The more complex circuit of Fig. 2G (f) is somewhat akin to a phase locked loop, in that it feeds back from its output an estimate of the current pulse-rate, and uses this estimate to adjust the detector blanking period. Thus Fig. 2G (d) employs a fixed blanking period (= 0.9 x minimum period), while Fig. 2G (f) employs an adaptive blanking period (=0.9 x estimated current period). To design such an adaptive loop, it is necessary to know the maximum rate at which the incoming pulse rate will change, which is determined by the bandwidth of the original modulating signal. Adaptive blanking cannot be used when the incoming pulse train possesses serious jitter, e.g. comes directly from a binary rate multiplier.

In the table below a comparison is made between the techniques represented by Fig. 2G (e) and Fig. 2G (f); that is, between the use of an adaptive threshold with fixed detector blanking, and the use of a fixed threshold with adaptive detector blanking. As before, a pulse of length T and of unit height, is mixed with wideband Gaussian noise. Only the pulse-period is varied. The 'optimum' threshold is that which gives equal probability of positive (false pulse) and negative (missed pulse) errors. Noise power = 4 units (c.f. filter comparison table earlier).

Pulse period	2T	3T	4T
Probability of error using optimum threshold*	.01	.09	.14
Probability of error using adaptive blanking**	.01	.01	.01

* Blanking constant at 1.8T ** Threshold constant at 0.36

Clearly adaptive blanking should be used wherever possible. Any adaptive process is likely to fail in the presence of very high noise levels, in the same way that FM is inferior to AM when noise exceeds a certain high threshold.

A full analysis of the performance of PFM detection has not been undertaken. Detection is a non-linear process which reduces the effects of noise at low noise levels. The effect of false detections, be they negative or positive, depends on the relationship between a pulse or event and the information being conveyed. It has already been argued that errors due to time displacements in detection are not generally significant, particularly if the modulation index is high. The matched filters mentioned earlier are optimum for the detection of single pulses in white noise. The adequate detection of a train of pulses may be approached via FM theory so that a band pass filter is chosen, or via detection theory so that a low pass (matched) filter is chosen and used in conjunction with detection blanking. The two approaches meet in the full sophistication of a phase locked loop. If additive noise is very low and the modulation index is not high then time displacements are the only error mechanism. In such cases very wideband filters should be used to increase the ratio of pulse-front slope to r.m.s. noise; choice of upper frequency limit is determined by phase distortion of the pulses during transmission.

At moderate noise powers, the probabilities of missed-pulse and illusionary-pulse errors will be low. The error pulses may be assumed to have a Poisson distribution. Detector thresholds may be adjusted to give equal probability of +ve and -ve errors, although this equality cannot be maintained over a range of noise power. The significance of miss-detections is affected by the use made of the detected pulses. If detection is followed by periodic counting, or by very-low-pass filtering, the mean error in the count or output voltage can be kept very small. The root mean square error rises with (at least) the square root of the error probability. The expected error modulus is about 20% less than the r.m.s. value.

[Note: For a Poisson process running for such a time that on average A events are generated, and where the value attributable to any event is +1 or -1 with equal probability:-

then variance of total value = A x 1, expected total value = 0, and

r.m.s. total value = $\sqrt{\text{variance}} = \sqrt{A}$.

Also it can be shown that

$$\text{expected modulus of total value} = e^{-A} \sum_{i=0}^{\infty} \left(\frac{2i}{A} + 1\right)\left(\frac{A^i}{i!\,2^i}\right)^2$$

$$\doteqdot 0.78\sqrt{A}, \text{ if } A > 6 \quad]$$

Experiments were carried out that confirmed this relationship between r.m.s. and mean modulus errors.

A detector was simulated that incorporated a matched filter, an optimum threshold and optimum blanking. It was fed with a uniform pulse train embedded in wideband near-Gaussian noise. Detection errors were very rare with post-filtering signal-to-noise power ratios of better than 5. Even with a signal-to-noise ratio of 1, the rate of detection errors did not exceed 4% of the incoming pulse-rate.

DIGITAL TECHNIQUES

Contents:

This long chapter is concerned with digital techniques for processing pulse-frequency-modulated information. The emphasis is upon evaluation; most of the techniques have been described before, but they have not been related to each other.

The three building blocks for most pulse-frequency digital circuits are the coincidence gate, the bidirectional counter and the frequency scaler. In the case of the first, the alternatives are asynchronous and synchronous modes of working. With the second the choice lies between synchronous, semi-synchronous and ripple designs. The main contenders for the third are designs based upon the binary rate multiplier and upon the binary rate divider. Sections 3.2, 3.3 and 3.4 consider these building blocks in turn. In the case of the last (scalers), an extensive analysis has been undertaken which shows that for all categories of error: binary rate divider designs are better than binary rate multiplier designs. Current practice is however wholly based on the latter. For a first reading it is suggested that section 3.4 be omitted except for the conclusions in part 3.4.4.

Sections 3.5 and 3.6 cover cancelling, frequency comparison and the two main methods of representing negative quantities.

Sections 3.7 and 3.8 contain descriptions and comparisons of techniques to convert frequency, and rate-of-change of frequency into number. It is shown that the widely used sampling methods are not always the best. A novel high speed frequency-to-number converter is described in part 3.7.1. A statistical analysis of the rounding errors in frequency and frequency-rate to number conversions, shows them to be greater than the corresponding errors in conventional analogue-to-digital conversion.

The final sections describe more complex circuits to handle non-linear functions and operations requiring a degree of prediction.

3.1 Pulse-numbers and pulse-rate circuits.

The 'digital' circuits to be described in the following sections handle only binary signals. The theory of the circuits is applicable to electrical, pneumatic, hydraulic and other forms of logic; but as all experimental work has been concentrated upon electronic devices, electronics terms will be used throughout the description.

All signals are constrained to one of two values, or states, denoted 'O' and '1'. Using positive logic for clarity, state 'O' is a low voltage, state '1' a higher voltage. The transition from one state to the other is assumed to take negligible time. A pulse is the sequence of states 'O', '1', 'O'; a negative pulse is the sequence '1', 'O', '1'. The duration of a pulse has no logical significance, (although it may be controlled to obtain reliable working of practical circuits). The sequence 'O', '1' is known variously as a 'rising edge', or an 'α transition', and the sequence '1', 'O' as a 'falling edge' or 'β' transition. The instant of occurrence of a pulse is usually taken to be that of its rising edge.

Binary pulse circuits have been developed over many years. [301-306] Those of interest here found application in early computers, and in direct digital analysers. They perform arithmetic and transfer operations using the convention: n pulses represent the integer n. Pulse circuits may be distinguished from other digital circuits by their following characteristics:

(i) Arithmetic and transfer operations are inseparable, processing is performed in transit.

(ii) Every pulse has the same value or weighting.

(iii) Information passes serially along single channels; there are no parallel transfers.

The maximum speed of pulse circuits is determined by so-called ripple-through considerations. Any change in the overall state of a digital system due to the passage of one pulse should be completed before the passage of another pulse. This time of passage, or settling time, is short with electronic circuits, because logic transitions take only a few nano-seconds and the pulses travel between logic units at the speed of light. In process control applications pulse repetition rates above 100 k Hz rarely occur, and such rates can easily be handled by currently available integrated circuits. In certain cases a circuit designer can obtain extra speed at the expense of circuit simplicity: it may be possible to reduce the effective number of stages through which a pulse must ripple.

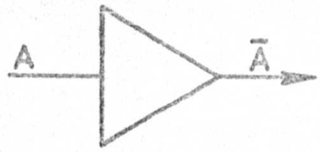
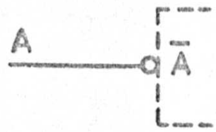
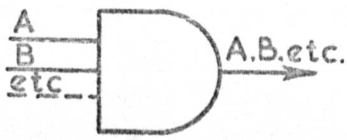
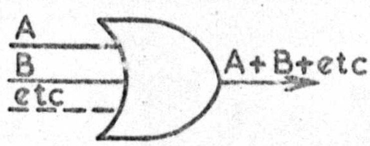
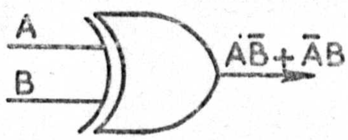
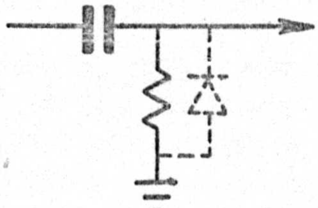
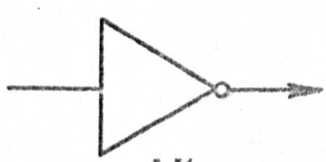
When a pulse is directed to a logic gate by two different but parallel routes there will be some uncertainty as to its arrival time. If the circuit contains loops round which pulses can travel, there is the possibility of instability and the generation of oscillations. These effects are known as problems of 'race', and can be avoided by careful design and layout, by the addition of extra gates redundant to the basic logic, or by the introduction of known delays in certain transmission paths,[307,308].

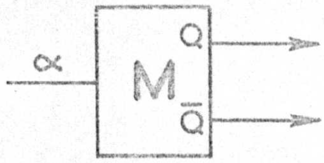
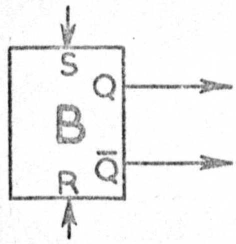
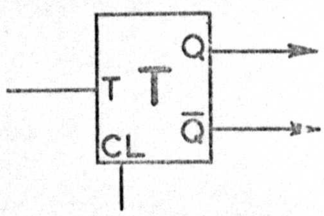
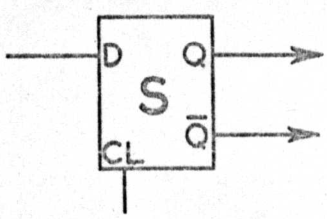
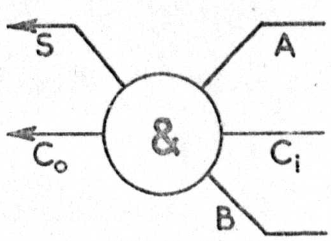
Other than as a constraint to avoid malfunction, speed or time does not enter explicitly into pulse-number circuits, but does do so in pulse-frequency circuits. Pulse-number computations are characterised by an initial (or 'reset') state, and a final or completion state. Pulse-frequency computations are continuous, without necessarily having any start or finish. A time scale is incorporated in pulse-frequency circuits by the supply of a 'clock' pulse-train of constant and known frequency.

Linear arithmetic operations are not time dependent; consequently the circuits for them are common to pulse-number and pulse-frequency systems. Addition, for example, is implemented by merging pulses from two lines into a train on a single line. Non-linear arithmetic (squaring, multiplication etc.), performed on the information carried by a pulse train, requires quite different circuitry for pulse-frequency encoding than for pulse-number encoding. Calculus operations involving time (differentiation and integration) can only be applied to pulse-frequency signals, but analogous operations may exist for pulse-number signals. For example, a counter fed by a pulse train acts as an integrator (of the pulse-frequency), or as an accumulator (of the pulse-number).

With both number and frequency encoding, sign presents special problems. The use of bipolar pulses requires ternary logic elements, and great practical difficulties ensue. The two practical methods of treating sign are (i) the use of a bias or offset so that all signals appear to be positive; (ii) the use of an auxiliary signal, a binary logic level, to indicate sign. The latter is usually the easier, especially now that bi-directional counters have become fairly cheap.

Pulse-number and pulse-rate circuits have much in common, and the latter may be thought of as a special case of the former (using an auxiliary signal, viz. the clock train). Circuits designed originally for pulse number work form the basis for the more recent pulse-frequency developments. Basic logic circuits, as found in all digital systems, are shown with their symbols on the following two pages.

FUNCTION	SYMBOL	NOTES
INVERSION (with buffering)		
INVERSION (as part of a more complex function)		
AND GATE		
OR GATE		
EXCLUSIVE OR GATE		
DIFFERENTIATING CIRCUIT		Responds only to rising edges (i.e. transitions)
SHAPER, BUFFER or SCHMIDT TRIGGER		Sharpens up pulse edges

FUNCTION	SYMBOL	NOTES
MONOSTABLE		A positive pulse of fixed duration is generated upon arrival of an ∝ transition at the input.
BISTABLE		S = 1 makes Q = 1 R = 1 makes Q = 0 R.S = 1 is forbidden R + S = 0, Q is frozen.
TOGGLE or DIVIDE-BY-2 CIRCUIT		If T = 0 then Q is frozen If T = 1 then Q inverts upon each ∝ transition of the clock line.
SHIFT REGISTER		Q takes the state of D upon each ∝ transition of the clock line
FULL ADDER		S is SUM A and B are INPUTS C_i is the CARRY IN C_o is the CARRY OUT.

It is beyond the purpose of this thesis to describe well-established digital circuits. Listed above are a number of circuits that often occur in pulse-frequency work, together with their symbols.

3.2 Synchronisation and Coincidence Circuits.

3.2.1 The problem of pulse overlap.

The addition (merging) or subtraction (cancelling) of two pulse trains is only meaningful if the pulses of the two trains do not overlap. Overlapping or coincident pulses can be handled in a number of ways: the simplest way is to ensure the pulses are very narrow. Given two randomly spaced pulse trains (whose mean rates represent two quantities), then the probability of overlap is a function of the mean mark-to-space ratios of the two trains (e.g. 1 microsecond pulses with a mean rate of 2 k Hz will overlap with probability 0.004). The probability of overlap and consequent circuit malfunction may be acceptably low. With evenly spaced pulse trains, although the same probabilities of overlap will apply, quite unacceptable particular cases can arise. For example, if both pulse trains have nearly the same repetition rate, there will periodically occur a situation wherein a whole string of successive pulses on the two lines are coincident. If the frequencies are 2000 Hz and 2001 Hz, and all pulses of duration 1 micro-second, then once every second 8 consecutive pulses on one line will over-lap pulses on the other, probability being 8 ÷ 2000 = 0.004 as with random spacing.

Two system strategies can be employed to avoid problems of overlap. One strategy is to use coincidence gates at every junction. Such gates inspect the incoming lines for overlapping pulses, and where these are found delay one pulse until the other has passed. The other strategy, called synchronous working, is to constrain all pulses to occupy specified slots in the continuum of time. A set of such slots is called a 'phase' of a clock. Other phases of the same clock will be sets of time slots interleaved with the first. The uniform spacing of any one set of slots is the reciprocal of the clock frequency. Note that if a signal pulse-train is constrained to a clock phase, some slots will be occupied by signal pulses, others will be vacant. Pulse trains synchronised to different phases of the same clock cannot overlap; they may therefore be safely added or subtracted.

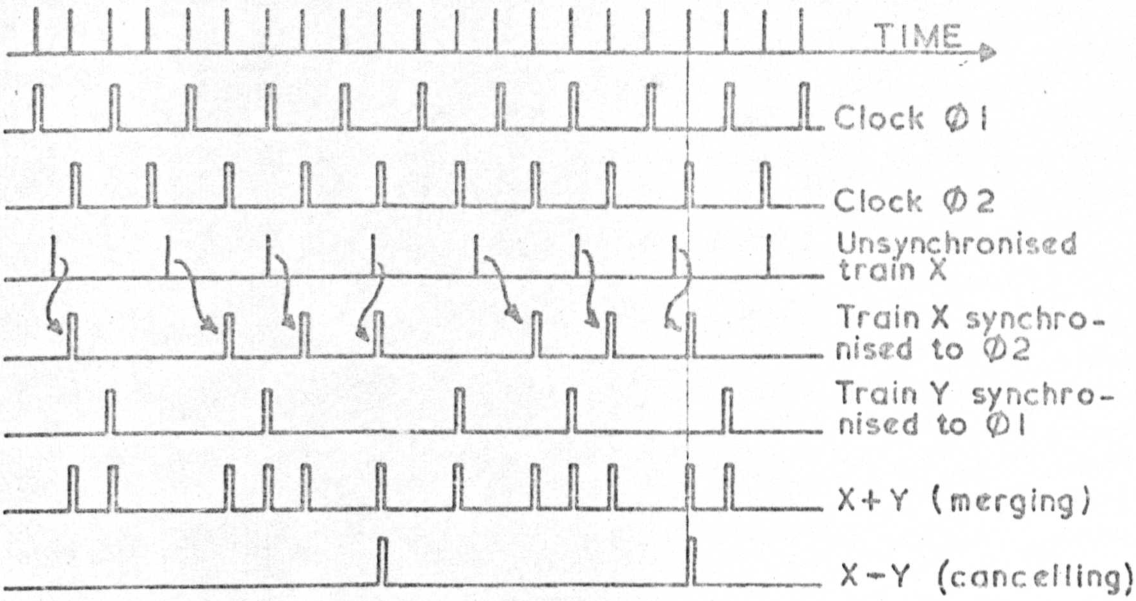
Figure 3A shows two phases of a clock, how an unsynchronised pulse train is synchronised to one of the phases, and how two synchronous trains are added and subtracted.

3.2.2 A coincidence gate.

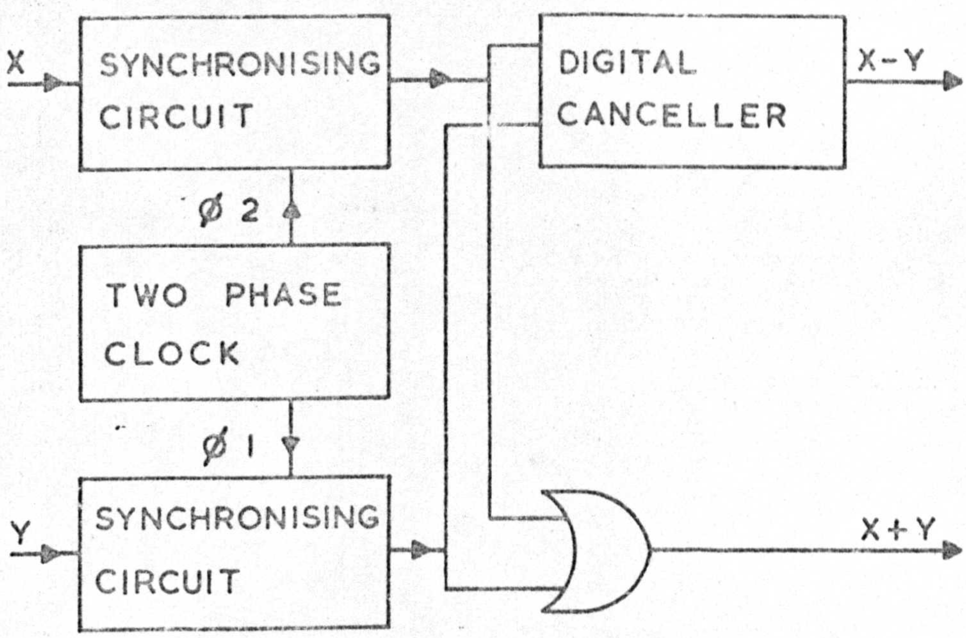
A coincidence gate can be compared to a policeman at a road junction, who observes the approach of vehicles on the two roads, and where a collision might ensue gives one vehicle priority and delays the other. As electrical pulses travel at the speed of light, their approach to a junction can only be observed if they are passed through some delaying section. A decision must be made, before the pulse leaves the delaying section, whether it is safe for it to do so, or whether the delay should be extended. Unless fairly complex circuitry is to be used, no pulse should be delayed so long that it could interfere with the next following pulse on its own line.

If a coincidence gate followed by a summing junction may be likened to a regulated road junction, a coincidence gate followed by a subtraction circuit has no such analogy.

A number of circuits have been described, e.g. Gruzdev & Karpov [309], wherein coincident pulses on two lines are made to cancel one another, the pulses having been shaped to give them equal heights and durations. Two such (analogue) cancelling circuits are shown in Figure 3B, in each case



(a) WAVEFORMS



(b) CIRCUIT

FIG. 3A CLOCK PHASES & SYNCHRONISATION

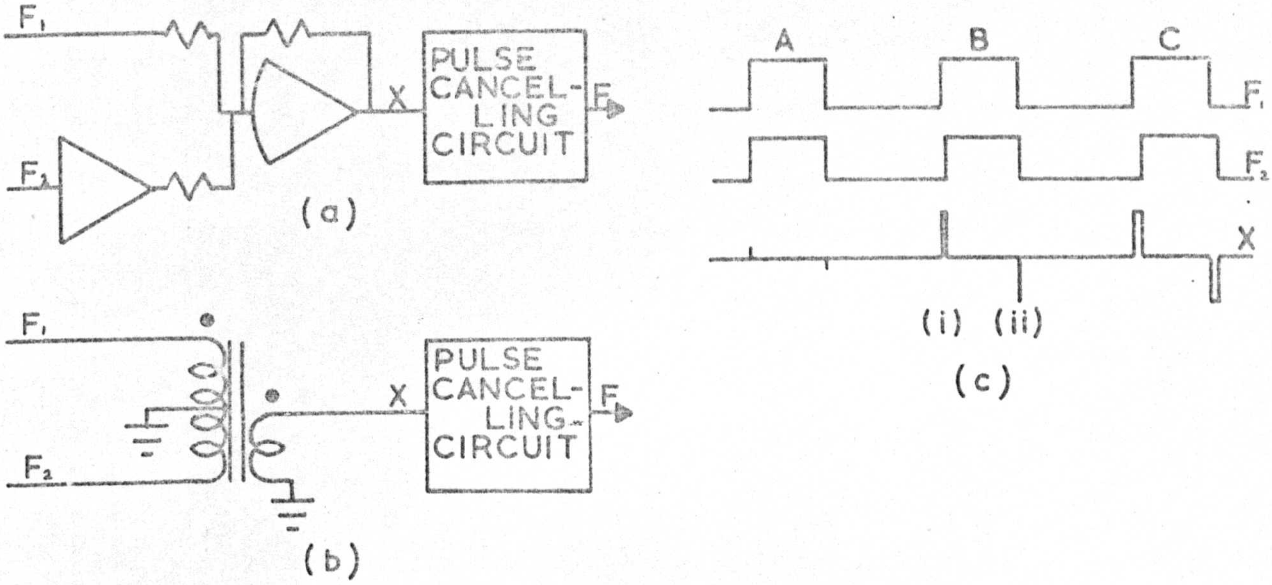


FIG. 3B SIMPLE COINCIDENCE CIRCUITS $F = F_1 - F_2$

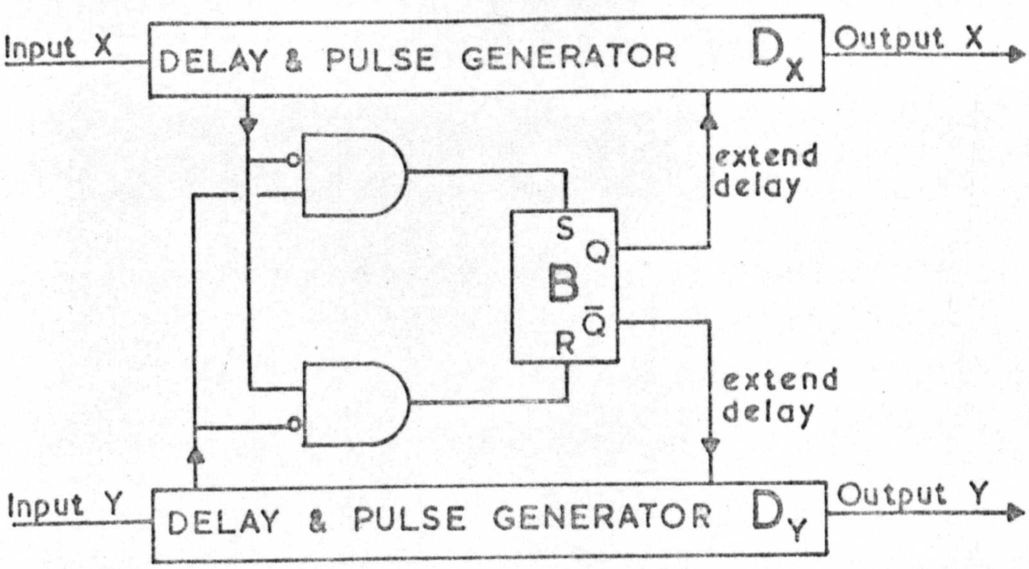


FIG. 3C COINCIDENCE CIRCUIT

preceding a digital cancelling circuit. The weakness of such circuits may be deduced from Fig. 3B(c). Absolutely simultaneous pulses of identical shape, such as pair A, are successfully cancelled by analogue means. Somewhat displaced pulses, such as pair C, emerge from the analogue circuit in a non-coincident form suitable for digital cancelling. However there remains the possibility of pulse pairs such as pair B that give an equivocal output after analogue processing. The pulses (i) and (ii) on the X line are narrow and not quite identical. The subsequent digital circuit may respond to one and not the other. While analogue cancelling reduces the likelihood of overlap malfunction, it only does so to the same extent as would the use of very narrow pulses, (i.e. so narrow as to just activate digital circuits reliably).

A small but finite uncertainty exists with all coincidence circuits, if a priority decision must be made within a specified time. As with other decisions, the smaller the precedence of one pulse over the other the less it matters which is allocated the priority. For pulses which arrive almost simultaneously, an arbitrary priority can be chosen, but some decision must be made.

The coincidence circuit developed is shown in Fig. 3C. Each input line feeds a delay and regeneration circuit, so that an output pulse is generated a known time after the α transition of an input pulse. These circuits are denoted D_x and D_y. The bistable B indicates the chosen priority. When B is set the Y line is favoured and the delay on the X line is extended; when B is reset, the inverse applies.

While D_x alone is active, B is reset; while D_y alone is active B is set. If both, or neither, D_x and D_y are active, the state of B is frozen.

In the case of clear priority, the prior pulse moves B to the state which favours its own line, - which state is maintained until the prior pulse has passed right through its delay circuit. In the case of uncertain priority, the bistable starts to change state and then is frozen. As long as it reverts to either set or reset condition before a D circuit delay has expired, there will be no malfunction. The probability of a bistable circuit not reverting to a stable state (in say 50 microseconds) is minute.

3.2.3 Circuits for synchronous working.

If synchronous working is adopted, the following circuits must be available:

(i) synchronising circuit

(ii) multiphase clock

(iii) phase shifting circuit to move a signal from one phase to another.

(i) A synchronising circuit without buffer storage must fail if more than one input pulse arrives in one clock interval; so the clock rate is always chosen high enough to avoid this possibility. Where this condition cannot be met, and a buffer is necessary, circuits discussed by Kemploy and Vernon [310] and by Vincent [311] may be used. Vincent shows that, for Poisson distributed random pulses, a divider makes a better buffer than a shift register.

Synchronising circuits can be divided into two classes, viz. simple, and with 'acknowledgement'. An example of each is shown in Figure 3D.

The simple form is represented by the circuit of Fig. 3D(a). Each input pulse sets a bistable. During arrival of each clock pulse (in the

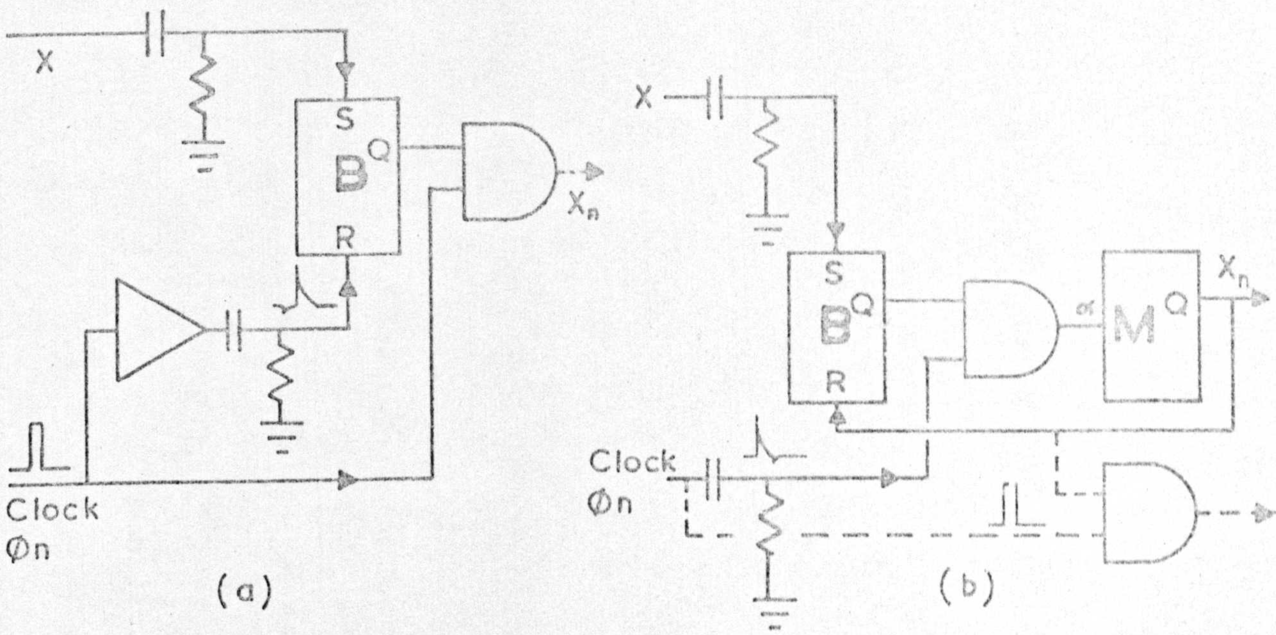


FIG. 3D SYNCHRONISING CIRCUITS

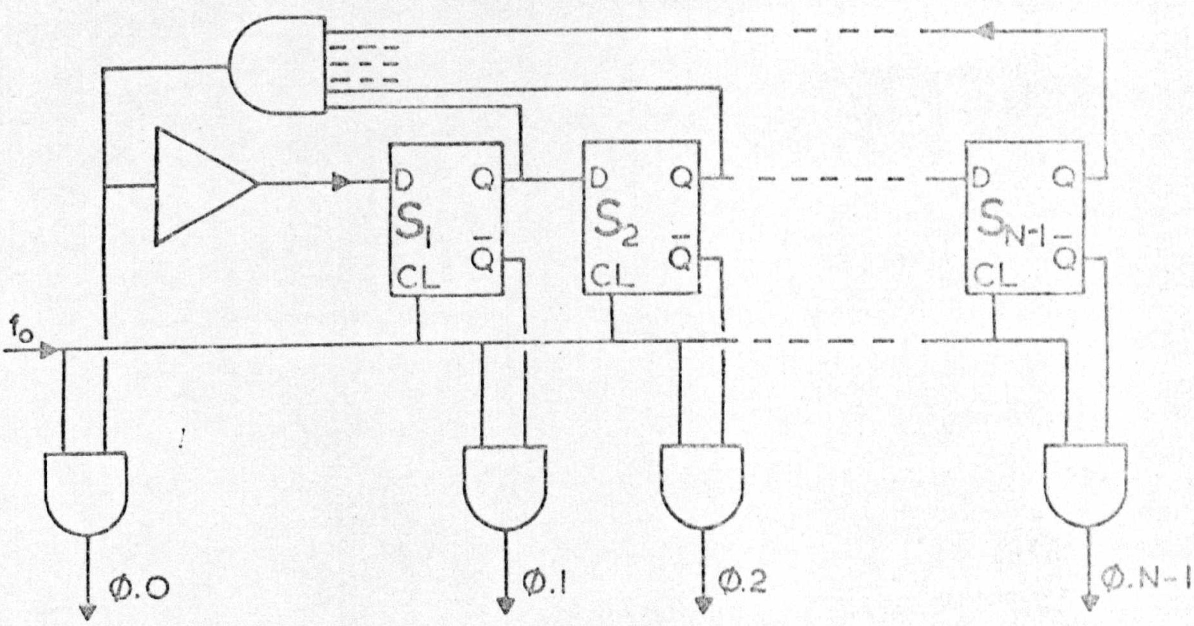


FIG. 3E MULTIPHASE CLOCK

figure, of clock phase n) the bistable state is read out, and upon fall of the clock pulse, the bistable is reset. Three modes of malfunction are possible with this circuit if the rise of an input pulse coincides with a clock pulse: the output pulse may be short, the output pulse may be absent or the input pulse may set the bistable twice and thus yield two output pulses.

The circuit of Figure 3D(b) contains a feedback link whereby the bistable is only reset when an output pulse is generated. Providing the clock rate is high enough, occurrence of an input pulse during a clock period guarantees the absence of another input pulse during the clock pulse at the end of that period. The bistable is thus only reset when there is no chance of a set signal arriving. Examination will show that the circuit functions correctly even when input and clock pulses coincide.

A number of versions of synchronisation circuits employing acknowledgement feedback exist, e.g. [312] which has a rather high cost.

(ii) A multi-phase clock circuit is shown in Figure 3E. This employs a ring counter and must be fed by an oscillator of frequency n times that of the desired clock rate. If n were very large, it would be advantageous to replace the ring counter by a binary counter with decoding of every second state. With the circuit of Fig. 3E, the input (oscillator) pulses should have a short fall time if spurious output spikes are to be avoided.

(iii) Phase shifting may be performed by simple synchronising circuits as represented by Fig. 3D(a). The feedback line of Fig. 3D(b) is not needed as there is no possibility of interrupt between pulses belonging to different phases of the same clock.

3.3 Counters.

Counters convert pulse-number (or 'serial') digital signals into a parallel coded form. Alternatively they integrate the frequency of a pulse train. As they overflow after a fixed number of input pulses, they can be used as frequency dividers. An up-counter increases its (coded) content by 1 for each input pulse. A bidirectional counter (BDC) has an auxiliary input: a logic line whose state determines the direction of counting. Counting up, the direction line is in state '1'; the count increments for each input pulse. Counting down, the direction line is in state '0'; the count decrements for each input pulse. To avoid malfunction, the direction line should not change during an input pulse.

Common electronic counter codes are binary, binary coded decimal (various versions) and cyclic. Mechanical counters are not constrained to binary working, and exist in decimal and other forms; they are however too slow for pulse-frequency applications. Gas discharge tubes have been used for counting, but their reliability, bulk and operating voltages make them unsuitable for incorporating in modern electronic systems. At present, transistorised electronic counters have no serious competitors for medium and high speed calculations.

While mechanical counters have been in use for over a century, and electrical ones for several decades, it is only in recent years that bidirectional counters have been commonly available. Integrated circuit BDCs reached the English market in Autumn 1969. The direct digital analysers constructed prior to about 1967 used combinations of up-counters to achieve bidirectional effects. Interest continues in these methods under the title of 'phase computers', Gaines and Joyce [313]; but whereas they have some use for pulse-number work, they are clumsy when used for continuous pulse-frequency circuits.

Counters may be divided into three types: ripple-through, synchronous and semi-synchronous. Each type can be realised in a number of ways, but the discussion here will be concentrated upon circuits using toggle bistables.

3.3.1 Counter Types.

A ripple-through up-counter can be made by cascading a number of toggle bistables. If any stage should 'overflow', i.e. make a $1 \rightarrow 0$ transition, the following stage is toggled, i.e. its state inverted. Consider, as an example, a 5-bit counter whose initial state represents the binary number 00111. On arrival of an input pulse, the first or least significant stage will invert, giving the count a momentary value of 00110. The next stage will now invert, as the first stage made a $1 \rightarrow 0$ transition. So the counter starting at 00111 passes through intermediate states 00110, 00100, 00000 to reach its final steady state 01000. The longer the counter, the longer the time taken for a change to ripple through all the stages, and the longer the counter may be in some false intermediate state. Many transitions do not involve the more significant stages of the counter, in which the change does not have to ripple through all the stages. For example, the change from 10000 to 10001 only affects one stage and is quickly completed. Should a counter be read, e.g. its contents copied into some other register, there is a possibility that it is in course of a transition and a false reading may result. In the case of 7 + 1 = 8 discussed above, the intermediate states of 6, 4 and 0 are grossly in error. The quicker the transitions, the lower the probability of such a reading error. This is discussed in greater detail at the end of this section.

When a simple ripple-through counter is modified for bidirectional working, a problem arises as follows. A change of logic level on the count direction line is not an actual count, yet is liable to be treated as such

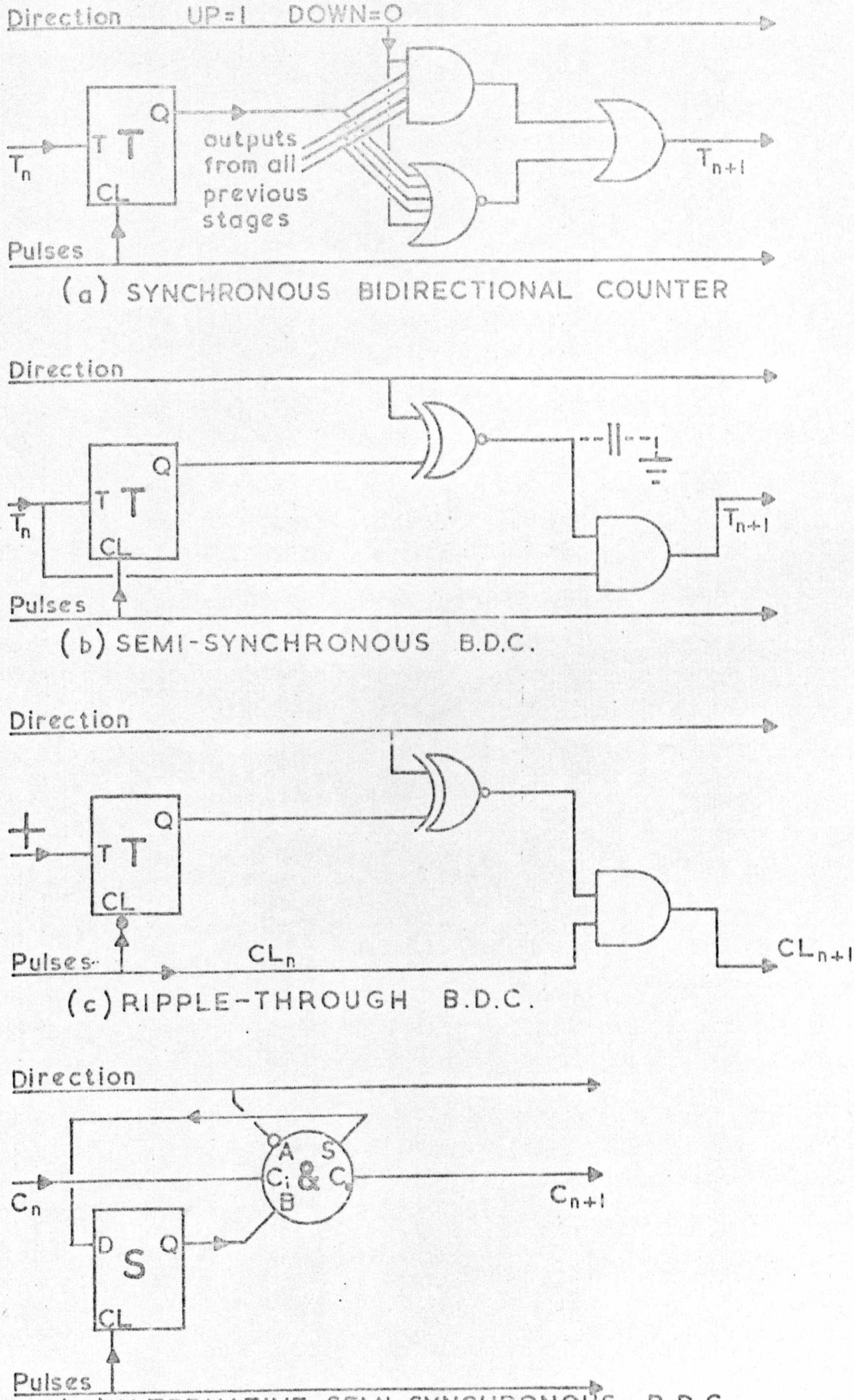


FIG. 3F TYPES OF BIDIRECTIONAL COUNTER

One stage is shown for each type.

by a ripple-through BDC. For this reason simple ripple-through BDCs are not used, but rather a design like that of Fig. 3F(c), which retains certain ripple-through features. Each stage is identical (i.e. modular construction); there is only one input to each stage; the maximum transition time for the whole counter is N x T, where T is the transition time for an individual gate or bistable.

Fig. 3F(a) shows the n^{th} stage of a synchronous BDC. This is fed from the n - 1 previous stages in such a way that it can make an immediate transition to its new state on arrival of an input pulse. For example when the counter changes from state 11111 to state 00000 (i.e. it overflows), every stage changes simultaneously. There is no rippling from stage to stage. Transition time is T, and a further time 2T is required before the conditions for the next transition are set up. Construction is no longer modular, and the interconnection of the stages becomes quite laborious for large counters.

Fig. 3F(b) shows one stage of a semi-synchronous BDC. Like the synchronous type, there is a common input or 'clock' line that ensures all stages change simultaneously. Like the ripple-through type, the construction is modular and each stage only communicates with one ahead and one behind itself. Counter transition takes only T seconds, but a further NT seconds are required for the logic levels (needed for a subsequent transition) to ripple-through the AND gates.

The properties of the three counter types are summarised in the following table:

Type of Counter	Synchr.	Semi-synchr.	Ripple
Diagram	Fig. 3F(a)	3F(b)	3F(c)
Modular construction ?	No	Yes	Yes
Inputs to N^{th} stage *	N + 1	3	2
Max. transition time	T	T	NT
Max. set-up time	2T	N.T	-
Maximum safe counting rate	1/3T	1/(N + 1)T	1/N.T +

* includes the direction-of-count signal.

\+ may be run faster during periods when no readout is required.

An operating constraint of some importance is the prevention of counter overflow. It is often necessary to clamp a counter at full (e.g. 111111) when counting up, and at empty (e.g. 000000) when counting down, rather than permit overflow. This can be achieved quite simply with the semi-synchronous design of Fig. 3F(b), by inverting the 'toggle' output T_{N+1} from the last stage and applying it as the 'toggle' input T_1 to the first stage.

3.3.2 Counter Errors.

Calculations now follow to establish the statistics of errors due to misreading the different types of counter while they are in transition.

Let F be the counting rate

N the counter length in bits

$M = 2^N$ the counter capacity

T the transition time for a gate or bistable

e the read error

then

$E = e/M$ is the normalised read error and $A = FT/M$ is a convenient constant in the ensuing analysis.

(i) For synchronous counters:

The probability of the counter being in transition at any instant is

$$P(Tr) = F.T \, , \tag{3-1}$$

while the probability of a transition affecting bits 1 to j only of the counter is

$$P(j) = F.T.2^{-j} \, . \tag{3-2}$$

Assumption: During a transition affecting bits 1 to j only of the counter, these bits have an arbitrary value, so that the number they represent is evenly distributed in the interval: 0 to 2^j-1. The consequent error is thus evenly distributed in the interval: -2^{j-1} to $+2^{j-1}$.

From the assumed boxcar distribution, the error variance during a '1 through j' transition is

$$V_e^2(j) = 2^{2j}/12 \, , \tag{3-3}$$

giving an overall variance of

$$V_e^2 = \sum_{j=1}^{N} P(j) \, . \, V_e^2(j) = F \, . \, T \, (2^{N+1} - 2)/12 \doteqdot FT \, . \, M/6$$

which after normalising and the substitution of constant 'A' yields

$$V_E^2 = V_e^2/M^2 = A/6 \, . \tag{3-4}$$

Using the assumption above, and writing $P(e_i;j)$ for the probability that $|e| > 2^i$ during a '1 through j' transition,

then

$$P(e_i;j) = 0 \quad \text{when} \quad j < i+2,$$

$$P(e_i;i+2) = \tfrac{1}{2}, \quad P(e_i;i+3) = \tfrac{3}{4}, \quad P(e_i;i+4) = \tfrac{7}{8}, \quad \text{etc.,}$$

so globally

$$\text{Prob.} \left(|e| > 2^i \right) = \sum_{j=1}^{N} P(e_i;j) \, . \, P(j) \tag{3-5}$$

$$= FT \left(\tfrac{1}{2} \, . \, 2^{-i-2} + \tfrac{3}{4} \, . \, 2^{-i-3} + \tfrac{7}{8} \, . \, 2^{-i-4} + \, . \, . \, . \right)$$

Substituting 'E' and defining $X = 2^{i-N}$, then equation (3-5) gives the series

$$\text{Prob.}\left(|E| > X \right) = \frac{A}{X} (1/8 + 3/32 + 7/128 + \ldots) \quad (3\text{-}6)$$

which truncates after $(-\log_2 X - 1)$ terms.

For $i \ll N$, equation (3-6) may be simplified to

$$\text{Prob.}\left(|E| > X \right) \doteqdot A/3X \quad (3\text{-}7)$$

Although equations (3-6) and (3-7) are only valid for $X = 2^{-i}$, where $i = 1,2,3,\ldots,N$, yet $\text{Prob.}\left(|E| > X \right)$ is a continuous smooth function of X, and may be interpolated.

(ii) For ripple counters:

The probability of the counter's j^{th} stage being in transition at any instant is

$$P(j) = 2^{1-j} . F . T \quad (3\text{-}8)$$

and in 50% of such cases the transition then ripples on to the $j+1^{th}$ stage. Thus the probability that the counter is in a transition state is

$$P(Tr) = \sum_{j=1}^{N} P(j) = 2 . F . T \quad (3\text{-}9)$$

Assumption: Any stage read during transition has an equal probability of having the right (new) or wrong (old) state.

The error during transition of the j^{th} stage has zero mean, and

$$\begin{aligned} e &= 0 \quad \text{with probability} = P(j)/4 \\ |e| &= 2^{j-1} \quad \text{"} \quad \text{"} \quad = P(j)/2 \\ |e| &= 2^{j} \quad \text{"} \quad \text{"} \quad = P(j)/4, \end{aligned} \quad (3\text{-}10)$$

giving a variance during this particular transition of

$$V_e^2(j) = \tfrac{1}{2} . 2^{2j-2} + \tfrac{1}{4} . 2^{2j} = 3 . 2^{2j-3} . \quad (3\text{-}11)$$

The global read error variance for a ripple counter, using (3-8) and (3-11), is

$$V_e^2 = \sum_{j=1}^{N} V_e^2(j) \cdot P(j) = 3FT \sum_{j=1}^{N} 2^{2j-3} \cdot 2^{1-j}$$

$$\doteq 3FT \cdot 2^{N-1} ,$$

and normalising

$$V_E^2 = V_e^2 / M^2 = 3A/2 . \tag{3-12}$$

As from (3-10)

$$\text{Prob.} \left(|e| = 2^j \right) = P(j)/4 + P(j+1)/2$$

so using (3-8)

$$\text{Prob.} \left(|e| \geqslant 2^j \right) = \tfrac{1}{4} \sum_{i=j}^{N} P(i) + \tfrac{1}{2} \sum_{i=j}^{N-1} P(i+1)$$

$$= 2 \cdot FT \cdot 2^{-j} - 1\tfrac{1}{2} \cdot FT \cdot 2^{-N} ; \tag{3-13}$$

after normalising, and substituting $X = 2^{j-N}$ and $A = FT \cdot 2^{-N}$,

$$\text{Prob.} \left(|E| \geqslant X \right) = A \left(\frac{2}{X} - 1\tfrac{1}{2} \right) . \tag{3-14}$$

This probability is not a continuous function of X, but only changes when X passes through a value 2^{-i}, $i=1,2,3,\ldots,N$.

So

$$\text{Prob.} \left(|E| \geqslant X \right) = A(2^i - 1\tfrac{1}{2}) \tag{3-15}$$

where

$$2^{-i} < X \leqslant 2^{-i-1} .$$

The following table illustrates the results just derived for the representative values :

T = 20 nanosec. for transistor-transistor gates

N = 10 bits for approx. 0.1% resolution.

F = 10 kHz maximum pulse rate for process control applications.

Counter type	Synchr.	Semi-synchr.	Ripple
P(Tr), probability of being in transition	2×10^{-4}		4×10^{-4}
Prob. $\left(\lvert E \rvert > 0.01 \right)$	6×10^{-6}		25×10^{-6}
Prob. $\left(\lvert E \rvert > 0.1 \right)$	0.5×10^{-6}		3×10^{-6}
Standard deviation of E	1.8×10^{-4}		6×10^{-4}
Maximum counting rate	17 MHz	$4\frac{1}{2}$ MHz	5 MHz

There are many situations where the errors shown above may be disregarded. The standard deviation of the normalised read error is less than 1/1024, the quantisation interval of a 10-bit counter.

Where the counter is used to control a multiplier, or feeds a display, or a digital-to-analogue converter, then momentary large read errors are of no consequence provided their overall variance is low. When the counter contents are read into a supervisory computer or alarm-limit scanner, momentary large errors may be inconvenient. From the table, with a synchronous type of counter, the probability of a read error exceeding 10% full scale is about 1/2,000,000, under representative operating conditions.

There are four methods of avoiding these large read errors. One is to address the counter only when it is known to be in a settled state. The second is to inhibit counting during reading, thereby increasing the probability of small errors due to 'lost' input pulses. The third is to test any reading for reasonableness (e.g. within 5% of last reading) and repeat it should it fail the test. The fourth method is to use a cyclic code, e.g. Gray code. Each method has its disadvantages. The first two require some form of signalling between counter and reading device. The last two require extra computational facilities in the reading device. Gray coded counters have been proposed, but are very complex.

There are counters that require no decoding at all, viz. ring counters, but these are prohibitively wasteful of bistable elements, and are difficult to design so that they do not overflow.

A bidirectional counter based on a conventional parallel adder has been described, Texas Instruments [314]. One stage of this is illustrated in Figure 3F(d). The adder performs the sum $A + B \rightarrow B$, where $A = 1$ for an increment, $A = 2^N - 1$ for a decrement. The sum is performed every input pulse. Despite its derivation from an adder, this BDC is quite similar to that of Figure 3F(b) based on toggle bistables, and may be classified as semi-synchronous.

For Fig. 3F(b), the logic equations are

$$T_{n+1} = (UQ_n + \bar{U}\bar{Q}_n)\,T_n \quad \text{and} \quad Q_n^{+} = \bar{Q}_n T_n + Q_n \bar{T}_n$$

For Fig. 3F(d), the equations are

$$C_{n+1} = \bar{U}Q_n + \bar{U}C_n + Q_n C_n \quad \text{and}$$

$$Q_n^{+} = \bar{U} \oplus C_n \oplus Q_n = \bar{U}\bar{C}_n\bar{Q}_n + UC_n\bar{Q}_n + U\bar{C}_n Q_n + \bar{U}C_n Q_n$$

In both cases U is DIRECTION of count, Q_n^{+} is state to which Q_n goes on arrival of a clock pulse.

The two designs are linked by the relationship

$$T_n = UC_n + \overline{UC}_n \tag{3-16}$$

3.3.3. Recommendation

Taking into account modularity of construction, adequacy of counting speed, size of read errors, ease of overflow protection and the cost of components: the semi-synchronous counter of form shown in Figure 3F(b) was chosen as being the most suitable for pulse frequency work.

3.4 Scaling Circuits.

3.4.1 General properties.

Scaling the frequency of a repetitive signal is more difficult than scaling the amplitude of an analogue variable, and is not so well studied as digital multiplication (the scaling of coded signals). Indeed, a circuit to change the frequency of a sinusoid will often differ from one to change the frequency of a square wave or of a pulse train.

Circuits have been described in part 2 of this thesis which will translate a signal from pulse-frequency into analogue form, and vice-versa. Using an analogue multiplier sandwiched between such circuits, it is possible to multiply the frequency of a pulse-train by a wide range of fixed factors.

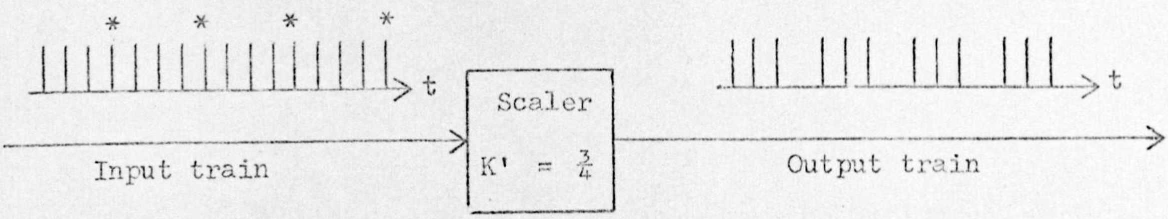
p.r.f. = f → [Demodulator] → [x K'] → [Modulator] → p.r.f. = K'f

Such a method of frequency scaling is not very accurate, or cheap; the process of demodulation introduces time lags; extra circuitry is required if K' is a variable presented in digital coded form.

Similarly, frequency scaling can be effected by converting the signal into coded digital form, passing it through a digital multiplier and then reconverting into pulse-frequency form. Greater stability is obtained, compared with the analogue method described above, but at much greater cost.

There exists a class of circuits that may be described as digital frequency scalers, which allow direct operation upon a pulse-train. Such a scaler comprises a gate and a memory, so connected as to remove a certain percentage of the pulses in a train. The ratio of output pulses to input pulses, K', is necessarily less than 1.

For example:



* Note: input pulses absent from output.

The factor K' is presented in digital coded form, and discussion will initially be limited to scalers using K' in binary code. If K' is specified as an N-bit binary fraction, it is convenient to introduce the N-bit binary integer K, where

$$K = K' . M \qquad \text{and} \qquad M = 2^N . \tag{3-17}$$

If the input pulses are evenly spaced and have repetition rate 'f', then the mean output repetition rate is

$$\bar{f}_{out} = K' . f = K . f_1 , \quad \text{where } f_1 = f/M . \tag{3-18}$$

The output pulse train is not usually an even one; the separation 'τ' of successive pulses may be greater or less than the mean output spacing '$\bar{\tau}$' (= $1/K . f_1$) . Whatever the pattern of output pulses, it will repeat after time:

$$T = 1/f_1 = M/f = K/\bar{f}_{out} \tag{3-19}$$

The unevenness of the output train may be measured in a variety of ways: different criteria of output jitter fitting different applications. Four criteria are proposed below.

(i) Variance of pulse spacing:

The normalised variance of the inter-pulse spacing is:

$$\sigma_{\tau}^{2} = \frac{1}{K} \sum_{j=0}^{K-1} \left(\frac{\tau_j}{\bar{\tau}} - 1 \right)^2 , \qquad (3\text{-}20)$$

where τ_j is time between the $j-1^{th}$ and j^{th} output pulses. This variance can be quite high, (up to $\frac{1}{8}$), so that an estimate of the frequency of a scaler output based on a single inter-pulse timing could be seriously in error. A better estimate of frequency is $\hat{f}_{out} = n / \tau_n$, where τ_n is the time taken for n output pulses to pass. A normalised n-period variance, $\sigma_{n\tau}^2$, can be defined in a way similar to σ_{τ}^2. If the errors in successive periods were uncorrelated, then $\sigma_{n\tau}^2 = \sigma_{\tau}^2 / n$. For actual digital scalers successive periods are correlated in a complex way, and the expression is only approximately satisfied. Because of pattern repetition after M input (i.e. K output) pulses,

$$\sigma_{n\tau}^2 = 0 \qquad \text{when } n = K, 2K, 3K, \text{ etc.}$$

(ii) Patterning and low frequency content.

An ideal scaler output, with its uniform pulse-spacing, would have a frequency spectrum consisting of lines at d.c. and at integer multiples of $\bar{f}_{out} = K \cdot f_1$. The d.c. component could be extracted using any low-pass filter with a cut-off frequency below $\bar{f}_{out}$. With a digital scaler the output spectrum is more complex, the output train possibly having components at all integer multiples of frequency f_1 (= f / M). A low-pass filter to remove all ripple would now need to have a cut-off below frequency f_1. Extracting the d.c. component of a pulse-train is the simplest way of demodulating its signal; so that reducing the cut-off frequency of the demodulation filter reduces the useful signal band-width. Thus passing a pulse-train through a digital scaler may corrupt the signal it carries.

It is therefore useful to measure the low-frequency ($< K . f_1$) content of a scaler output, and rate the scaler design as good if such content is small.

(iii) Pulse-position error:

The output of a digital scaler, and that of an ideal scaler can be compared. The pulses of the former will in general occur slightly before or after those of the latter, an effect known as pulse jitter. This is shown in Fig. 3G. Where the time displacement of the actual output pulse from its ideal position is denoted $x . \bar{\tau}$. Then the variance: σ_x^2 of x is one measure of jitter, while the maximum: x_{max} of x is another. It is the size of x_{max} which determines the length of queueing buffers required to remove jitter effects in frequency comparators.

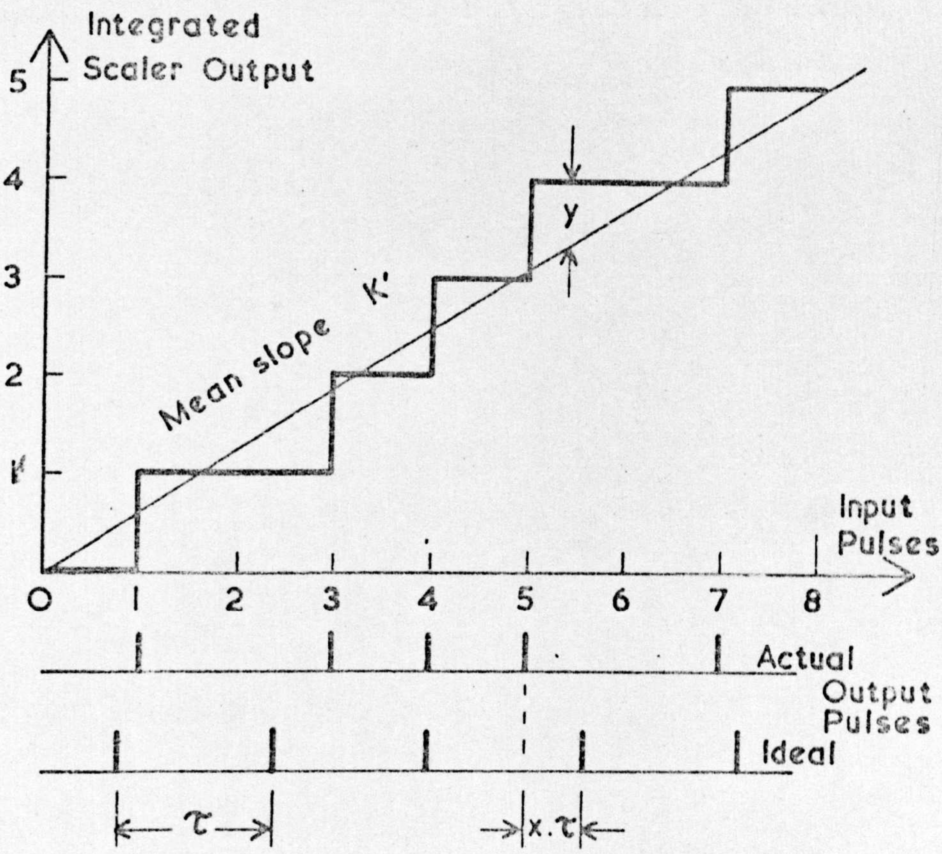


Fig. 3G. STAIRCASE AND PULSE POSITION ERRORS.

(iv) Staircase or cumulative error:

In direct digital analysers, interpolators and many pulse-frequency devices, a pulse train is fed via a scaler into an up-counter. When the input pulse train is uniform, the contents of the counter when plotted against time will give a staircase graph. Time is not implicit in the staircase relationship, which can be obtained by plotting counter content against input pulse number. Such a plot is shown in Fig. 3G.

An ideal scaler would yield an even staircase characteristic, symmetrically disposed about a ramp of slope: K'. If deviation of the staircase from the ramp is denoted 'y', then for the ideal scaler $|y| < \frac{1}{2}$ step. For a digital scaler the staircase is generally uneven, although each step is still 1 unit. Two measures of cumulative error are appropriate: the variance: σ_y^2 of y and the maximum: y_{max} of y. The latter measure assumes the staircase and ramp are so disposed as to make $y_{average} = 0$.

Digital scalers of size N bits have M internal states. The staircase obtained varies with the internal state of the scaler prior to arrival of the first input pulse; the staircase repeats its pattern after each M input pulses, (i.e. K output pulses). The average deviation from the ramp: $\bar{y}$ is some function of initial state, but as a form of symmetry is always present:

$$\text{when } \bar{y}=0, \text{ then} \quad y_{min} = y_{max} \; (=y_m) \quad ,$$

$$\text{otherwise} \quad -y_m \leqslant \bar{y} \leqslant y_m \tag{3-21}$$

$$-2y_m \leqslant y_{min} \leqslant 0 \leqslant y_{max} \leqslant 2y_m \, .$$

The staircase error 'y' and the pulse position error 'x' (defined above) are related by the staircase geometry. Defining 'y^+' as the value of y immediately following the output pulse whose position error is x, then

$$y^+ = 0.5 + x \quad \text{and} \quad \sigma_y^2 = \sigma_x^2 + 1/12 . \qquad (3\text{-}22)$$

For an ideal scaler

$$x = 0 \quad \text{and so} \quad \sigma_x^2 = 0 \quad ,$$

giving

$$y_m = 0.5 \quad \text{and} \quad \sigma_y^2 = 1/12 .$$

Of the various jitter criteria introduced in (i) above, only the maximum deviation y_m has received attention in the literature, and then only in relation to a particular scaler design.

3.4.2 The binary rate multiplier (BRM).

The binary rate multiplier has been known for over 15 years. Early descriptions of its principle and use were given by Lundh [301], Nicola [303], Arnstein et al [315] and Leonhard [305] among others. Interest has continued at a moderate level, and certain properties have apparently been rediscovered by different workers.

A BRM scales a pulse-train by first breaking it down into a number of interleaved pulse-trains, next passing each of these through a gate, finally recombining the gated signals. Consider an N-bit BRM which is fed by an input pulse-train of repetition rate: 'f'. $M = 2^N$ is the capacity of the BRM's internal register. Incoming pulses are distributed between N channels as follows. Channel 1 receives every 2nd pulse. Channel j receives every 2^jth pulse. Channel N receives every Mth pulse. The remaining pulses (also every Mth) go to an overflow channel. Pulse-frequency in the j^{th} channel prior to gating is $f \cdot 2^{-j}$.

The gate in the j^{th} channel is controlled by a logic signal denoted 'k_{N-j+1}' which has values 0 or 1. So after gating the pulse-frequency in the channel is $f \cdot 2^{-j} \cdot k_{N-j+1}$.

On recombination of the channels to give the output pulse-train, the mean output pulse-frequency is

$$\bar{f}_{out} = \sum_{j=1}^{N} f \cdot 2^{-j} \cdot k_{N-j+1} = f_1 \sum_{j=1}^{N} 2^{N-j} k_{N-j+1} = f_1 K \qquad (3\text{-}23)$$

where $f_1 = f / M$ and K is a binary integer.

One circuit for a BRM is shown in Fig. 3H(a). The input pulses enter an N-bit up-counter. The non-carry transitions of each stage of this counter trigger pulses into the corresponding channels. Note how the least significant bit of the counter toggles the most often and thus triggers the highest frequency pulse train in channel 1. Channel 1 therefore has the heaviest weighting (from a pulse-frequency point of view), and is consequently gated by the most significant bit (k_N) of the control register containing K.

Fig. 3H(b) shows an alternative circuit for a BRM.

Fig. 3H(c) shows pulse waveforms for a BRM (that is 4 bits long) for various multiplication factors K. Note how the output spacing is not even for some of the factors.

(i) Variance of BRM pulse spacing:

For M successive input pulses, there will be K output pulses, whose mean spacing is $\bar{\tau} = 1 / (f_1 K)$ where $f_1 = f / M$. Due to pulse-jitter, L of the output spacings will be too long (length 'τ_L') and S will be too short

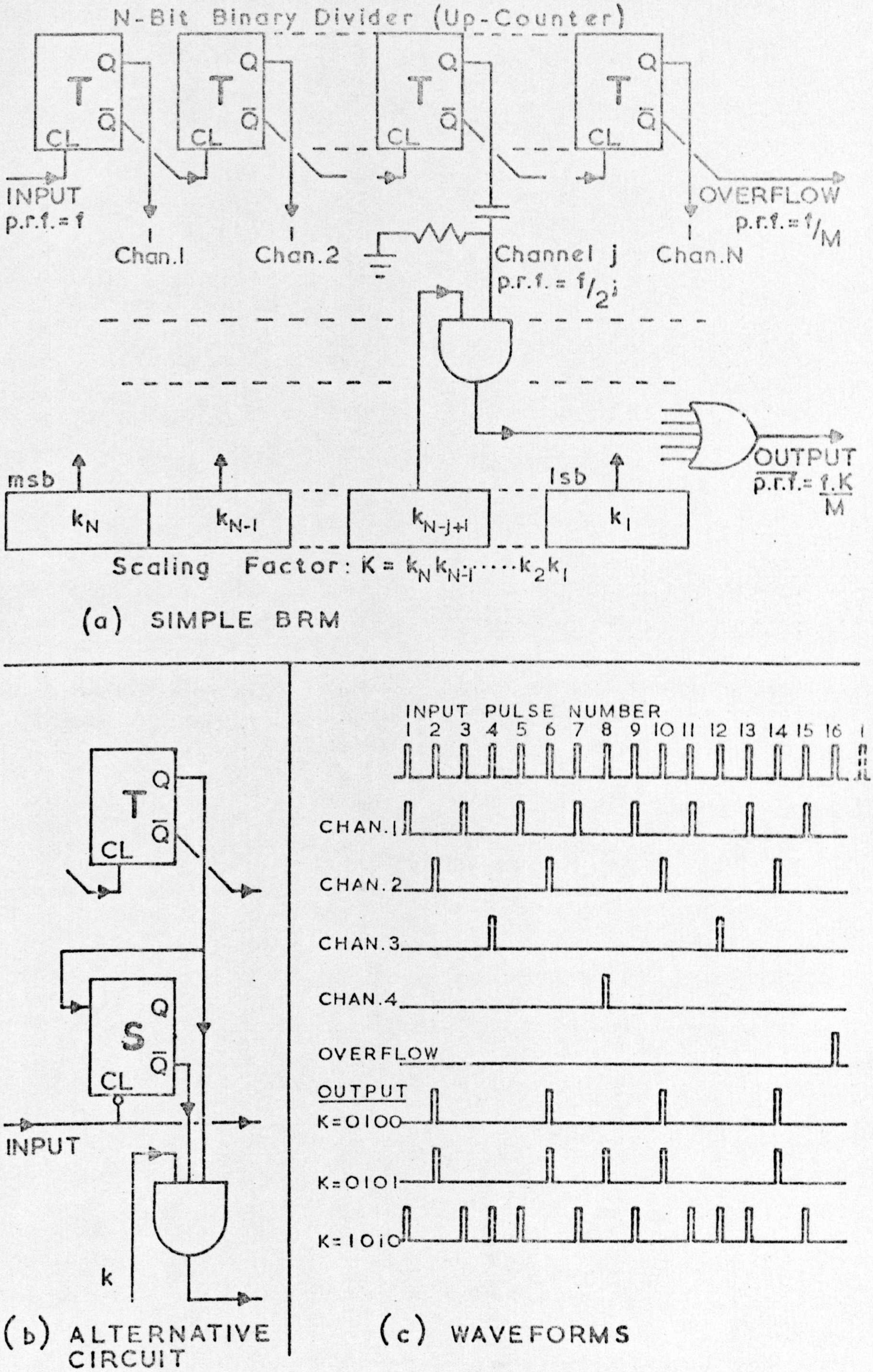


FIG. 3H BINARY RATE MULTIPLIER — BRM

(length 'τ_S'). Both τ_L and τ_S are necessarily binary multiples of the input spacings $1/f$.

Writing the scaling factor in mantissa-exponent form:

$$K/M = A \,.\, 2^{-B}, \text{ where } \tfrac{1}{2} \leqslant A < 1 \text{ and B is integer}$$

then:

$$\tau_S = 2^B/f \quad \text{and} \quad \tau_L = 2 \,.\, \tau_S ,$$

$$L + S = K = A \,.\, 2^{N-B} ,$$

$$L \,.\, \tau_L + S \,.\, \tau_S = 2^N/f. \qquad (3\text{-}24)$$

Solving:

$$S = (2 \,.\, A-1)\, 2^{N-B} \quad \text{and} \quad L = (1-A)\, 2^{N-B} \qquad (3\text{-}25)$$

The normalised variance of B.R.M. output period is then:

$$\sigma_\tau^2 = (L(\tau_L - \bar{\bar{\tau}})^2 + S(\tau_S - \bar{\bar{\tau}})^2)/K \,.\, \bar{\bar{\tau}}^2$$

$$= 3A - 2A^2 - 1. \qquad (3\text{-}26)$$

Equation (3-26) is illustrated by the upper curve of Fig. 3I. The period variance has maxima of $\sigma_\tau^2 = \frac{1}{8}$ at $A = \frac{3}{4}$ (i.e. $K/M = \frac{3}{4}, \frac{3}{8}$, etc.), while for $K/M = \frac{1}{2}. \frac{1}{4}, \frac{1}{8}$, etc. the output period has no variance.

<u>(ii)</u> Patterning and low-frequency content of B.R.M. output:
The spectrum of a pulse train depends slightly upon the shape of pulses, so that to simplify comparisons it is convenient to assume very narrow impulses of fixed 'weight' or area 'W'. A uniformly spaced chain of such pulses, i(t), of rate f_i, may be expressed by the expansion:

$$i(t) = W \,.\, f_i \left(1+2 \sum_{n=1}^{\infty} \cos\left[2\pi n f_i\, (t-t_i)\right]\right)$$

given that a pulse occurs at $t=t_i$. $\qquad (3\text{-}27)$

The output of the j^{th} channel of a B.R.M. is thus:

$$O_j(t) = k_{N-j+1} \cdot \frac{W \cdot f}{2^j} \left(1+2 \sum_{n=1}^{\infty} \cos\left[2\pi \frac{nf}{2^j} (t - \frac{2^{j-1}}{f})\right]\right)$$

and after substituting $i = N-j+1$ and $f_1 = f/M$

$$O_{N-i+1}(t) = k_i \cdot 2^{i-1} \cdot Wf_1 \left(1 + 2 \sum_{n=1}^{\infty} \cos\left[2\pi n(2^{i-1} f_1 \; t - \tfrac{1}{2})\right] \right)$$

So the total B.R.M. output can be written:

$$O(t) = \sum_{i=N}^{1} O_{n-i+1}(t)$$

$$= W \cdot f_1 \cdot K + 2\, Wf_1 \sum_{i=1}^{N} \left(k_i \, 2^{i-1} \sum_{n=1}^{\infty} \cos\left[2\pi n \; (2^{i-1} f_1 t - \tfrac{1}{2})\right] \right) \qquad (3\text{-}28)$$

The Fourier expansion (3-28) of a B.R.M. output may contain components at all harmonics of frequency $f_1 = f/M$,

i.e.

$$O(t) = Wf_1 \; K + 2 \; Wf_1 \sum_{m=1}^{\infty} a_m \cos\left[2\pi m f_1 t\right] \quad . \qquad (3\text{-}29)$$

As expected the d.c. component of the output (Wf_1K) is the pulse area times the mean output frequency (f_1K). The coefficients a_m in the expansion (3-29) are obtained by collecting suitable terms in expression (3-28), and are somewhat complex functions of K and m.

Writing:

$$\text{residue } (m \text{ modulo } M) = m_N \; m_{N-1} \; \cdots \; m_2 \; m_1$$

$$K = k_N \; k_{N-1} \; \cdots \; k_2 \; k_1$$

then

$$a_m = \sum_{i=1}^{N} x_i \, k_i \, 2^{i-1} \qquad (3\text{-}30)$$

where $x_i = 0$ if $m_j = 1$ for any $j < i$,

$x_i = 1$ if $m_j = 0$ for all $j < i$ and $m_i = 1$,

$x_i = -1$ if $m_j = 0$ for all $j < i$ and $m_i = 0$.

A representative BRM output spectrum envelope is shown in Fig. 3J.

For the example shown: $f = 1024$ Hz, $N = 10$ bits, $W = 1$ volt-second,

$\therefore f_1 = 1$ Hz and $Wf_1 = 1$ volt.

The scaling factor is 43/1024, i.e. $K = 43$.

Note: (a) d.c. component = 43 volts

(b) component at frequency 1 Hz is -2 volts ($a_1 = -1$)

(c) component at frequency 32 Hz is -42 volts ($a_{32} = -21$)

(d) component at frequency 43 Hz (= Kf_1, the mean output frequency) is only -2 volts ($a_{43} = -1$).

So to reduce all a.c. components to an amplitude less than 1 volt (the 'resolution' of the d.c. component) requires a filter with a cut-off frequency about ½ Hz, whose characteristic falls off at at least 6 dB per octave. The output 'power' is not concentrated at the mean output frequency of 43 Hz; on the contrary the power at this frequency is only 2 volts^2 compared with over 800 volts^2 at the lower frequency 32 Hz.

(iii) Pulse position errors of a BRM:

Because short output pulse spacings with a BRM consist of divided long output pulse spacings, the short (τ_S) spacing occur in multiples of two. This has the effect of severely displacing certain of the output pulses. Numerical values for these displacement are obtained by using equation (3-22) to translate staircase errors (y^+) into pulse-position errors. On the basis of section (iv) below, Figure 3K was developed, which also serves to illustrate pulse-position errors.

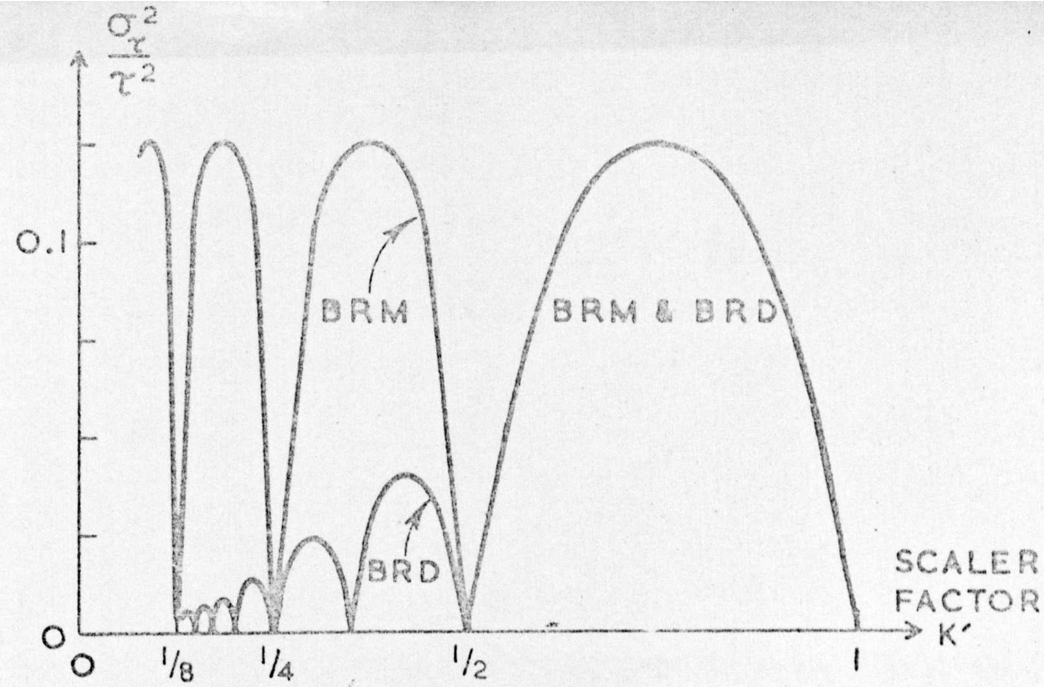
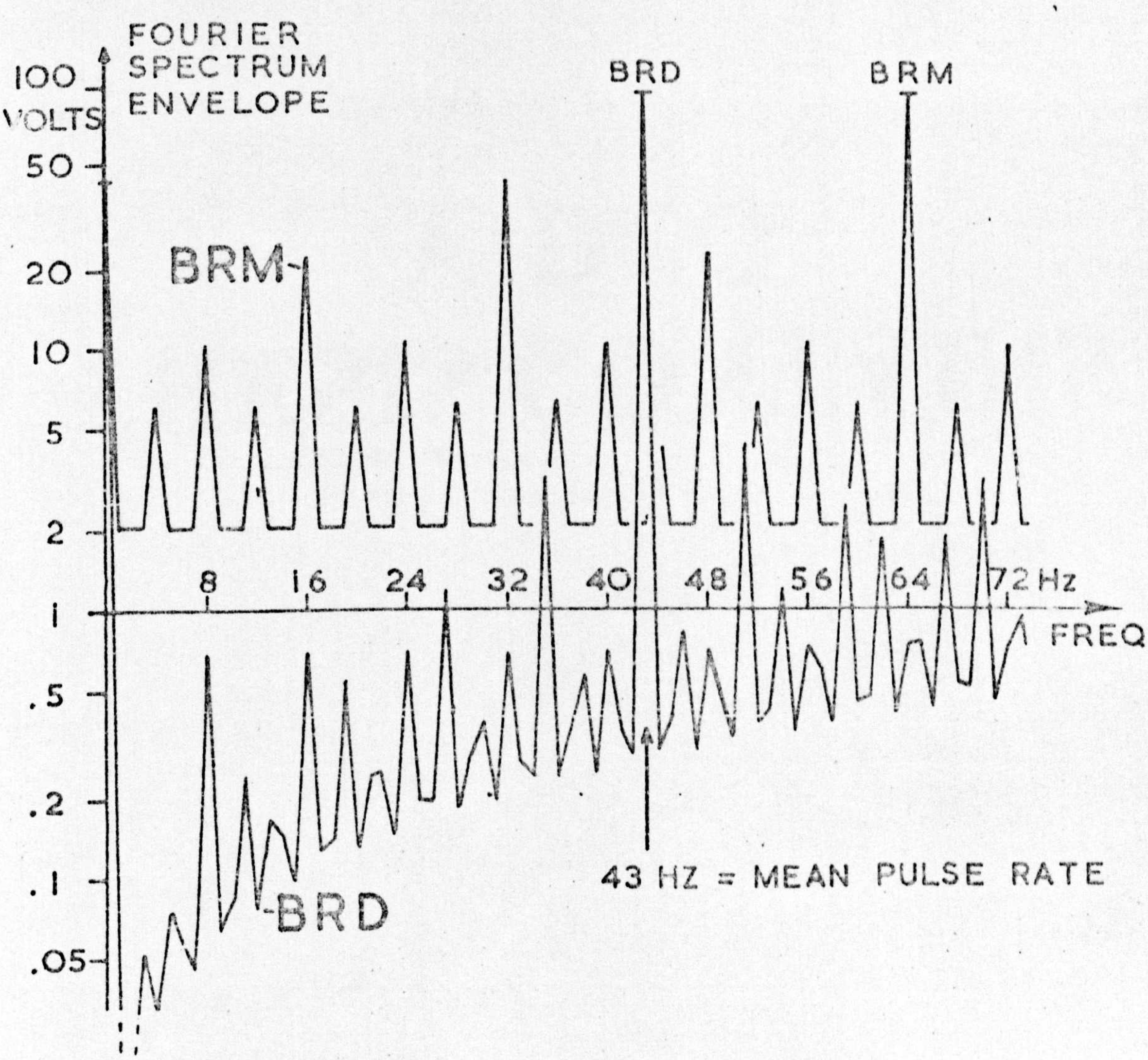
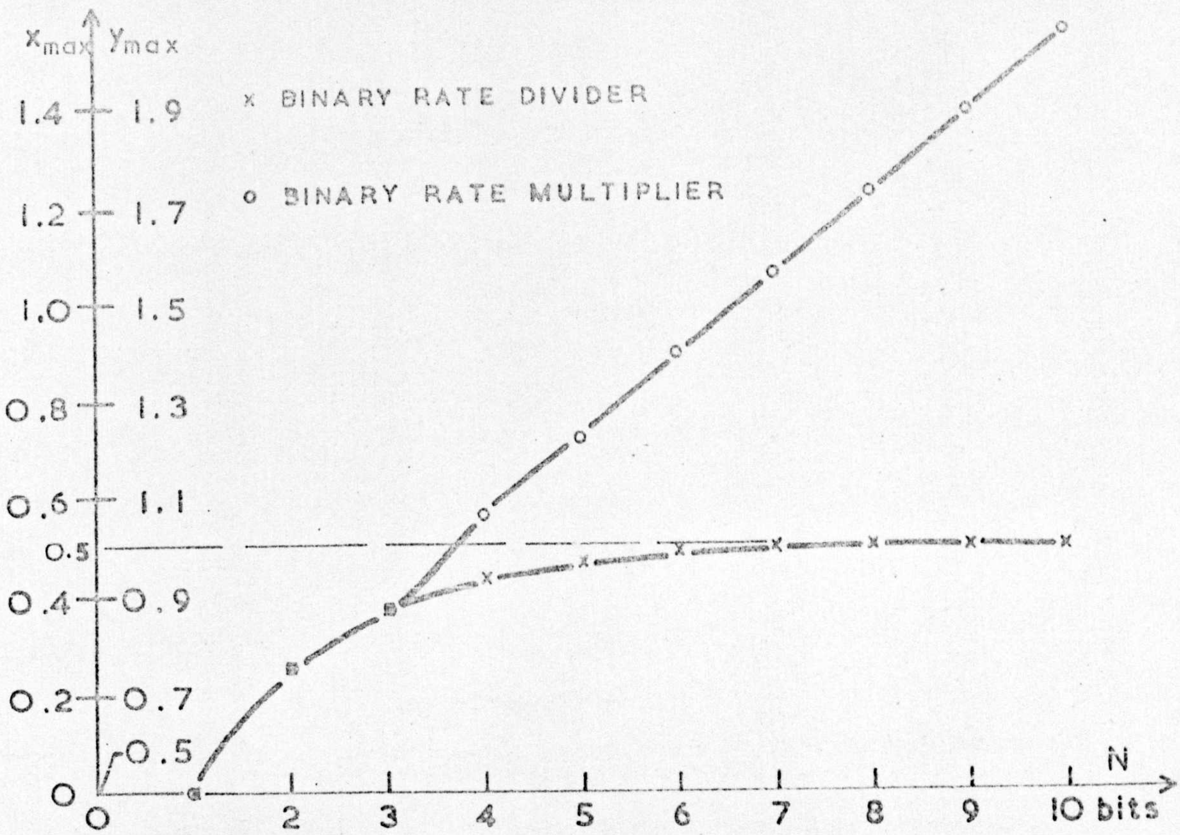


Fig. 3I. VARIANCE OF SCALER OUTPUT PERIOD.

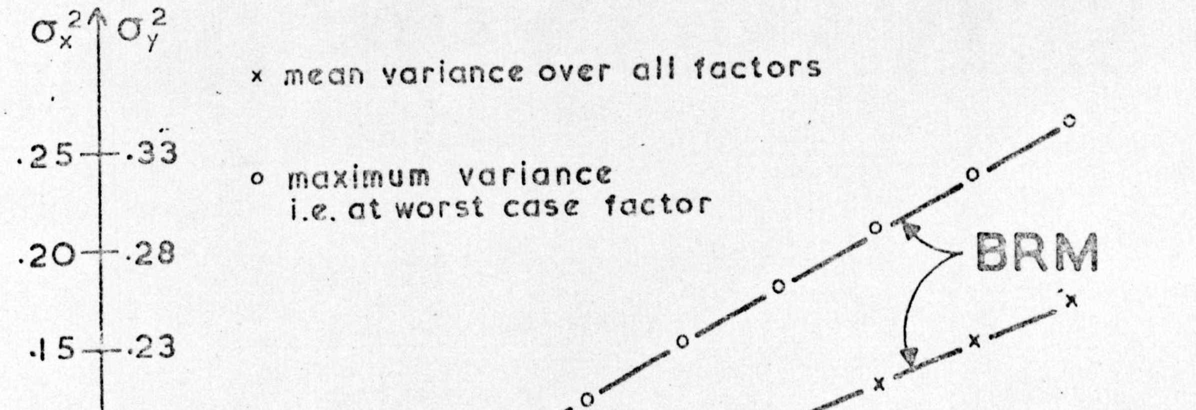


SCALE FACTOR = 43/1024 INPUT FREQ. = 1024 Hz
PULSE AREA = 1 VOLT-SECOND

Fig. 3J. FREQUENCY ANALYSIS OF SCALER OUTPUT.



(a) MAXIMUM ERRORS.



(b) VARIANCE OF ERRORS.

Fig. 3K. ERRORS AS A FUNCTION OF SCALER LENGTH N.
(X and Y defined by Fig. 3G.)

(iv) Staircase errors of a BRM.

The cumulative or staircase errors of a BRM output have been studied by Yang, 1960 [316], by Arnstein et al, 1964 [315], by Moshos, 1965 [317] and by Dunworth and Roche, 1969 [318]. These workers have deduced a general expression for y^+ and y^- (the staircase errors just after and just prior to an output pulse), in terms of the scaling factor K, the scaler length N, the state A of the scaler's N-bit internal register, and the initial value A_o of A. (The number of input pulses is $(A-A_o)$ modulo M). For $A_o = 0$, the mean cumulative error over M input pulses is zero. In these circumstances, the cumulative error after the A^{th} input pulse is: [316]

$$y^+ = \sum_{j=1}^{N} a_j \left(\tfrac{1}{2} k_{N-j+1} - \tfrac{1}{4} \sum_{i=1}^{N-j} k_i / 2^{N-j-i} \right) \qquad (3\text{-}31)$$

where $A = a_N \, a_{N-1} \, \ldots \, a_2 \, a_1$.

If y^+ is maximised with respect to number of input pulses (here equal to A) and the factor K/M, the value y_m as defined in equation (3-21) can be shown to be: [315-318]

$$y_m = 7/18 + N/6 + (-1)^N/(9 \,.\, 2^N) \qquad (3\text{-}32)$$

$$\doteqdot 7/18 + N/6 ,$$

$$\text{at } K/M \doteqdot 2/3 \text{ or } 5/6 .$$

Further maximising with respect to the initial state A_o gives a worst case staircase error of: (see 3-21))

$$y_{max} \doteqdot 7/9 + N/3 ,$$

$$\text{as } \bar{y} = y_m . \qquad (3\text{-}33)$$

y_m is plotted against N in Figure 3K(a), whose axis is labelled for both y_m and for pulse position error, x.

For $A_o = 0$ and therefore mean cumulative error $\bar{y} = 0$, the variance of y over M input (and hence K output) pulses was calculated for each value of K in the range 1 to M-1. The maximum variance occurs for those same values of K that give maximum y_m. This worst case variance is plotted in Fig. 3K(b), from which may be deduced the relation

$$(\sigma_y^2)_{worst} = 0.028\,N + 0.064\,, \tag{3-34}$$

at $\quad K/M = 2/3$ or $5/6$

when $\quad \bar{y} = 0$, and $N > 3$.

A more general criterion is the global variance over all factors K which is also shown for a BRM in Figure 3K(b), and which may be expressed by

$$(\sigma_y^2)_{mean} = 0.020\,N + 0.049 \tag{3-35}$$

when $\quad \bar{y} = 0$, and $N > 3$.

Both equations (3-34) and (3-35) were obtained by curve fitting to a set of computed results for N + 1, 2, ..., 9. No closed form expression has been discovered for the variance of the cumulative error, comparable to equation (3-32) for the maximum value of the cumulative error.

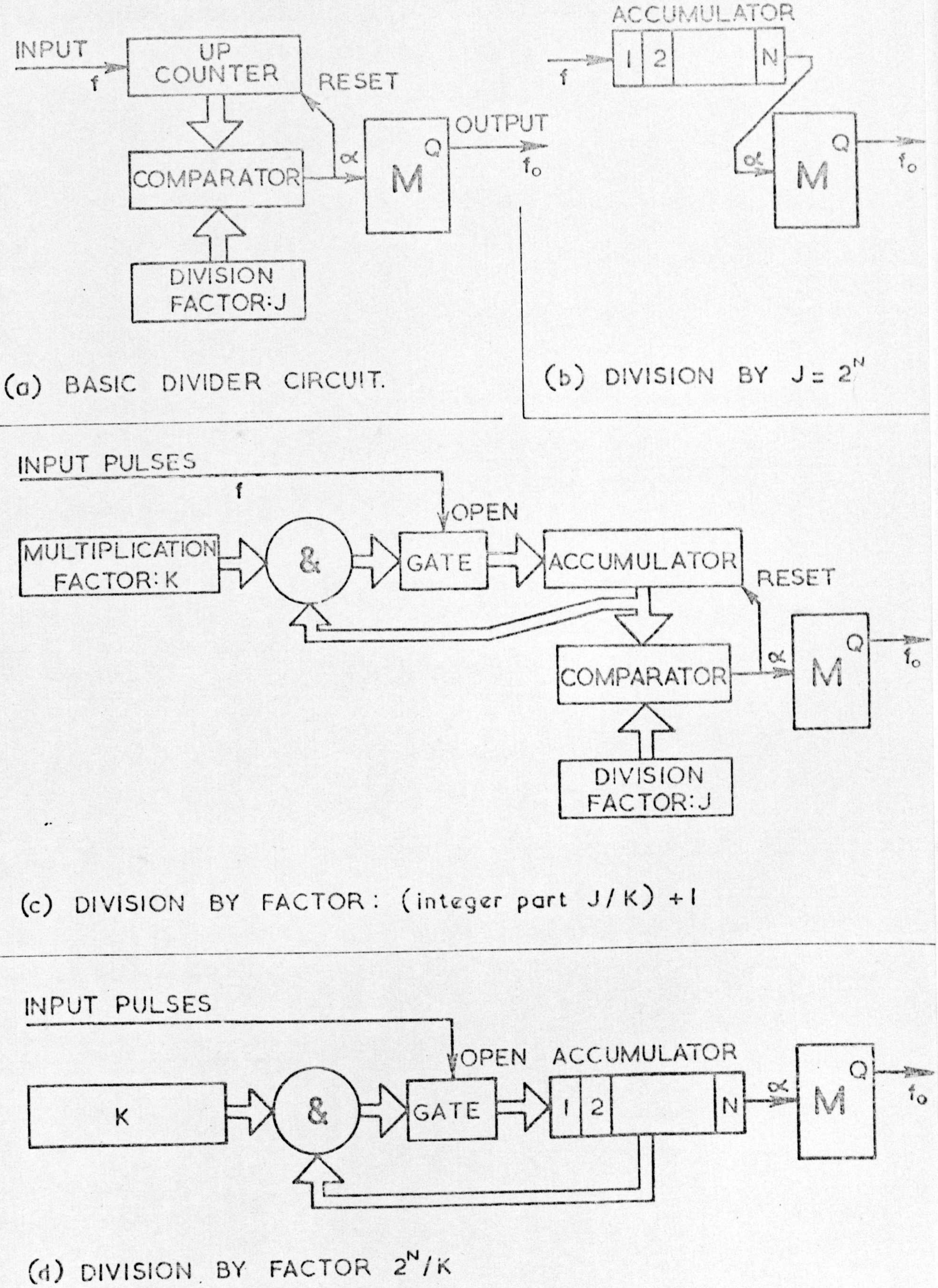
3.4.3 The binary rate divider (BRD).

A conventional circuit to select every J^{th} pulse from a pulse train is shown in Fig. 3L(a). Each incoming pulse causes 1 to be added to an accumulator. When the accumulator's contents reach J (as detected by the comparator), it is reset to zero and an output pulse is generated.

On arrival of input pulse: $1+A \rightarrow A$

If $A = J$ ....: $0 \rightarrow A$ and output pulse generated. (3-36)

This divider has a resolution of $1/J$; so to obtain fine resolution a frequency scaler based on division must reduce the frequency by a large factor. This reduction greatly reduces the usefulness of simple dividers in pulse circuits. Martin [320] has considered the properties of this basic divider when used as an operational element.



(a) BASIC DIVIDER CIRCUIT.

(b) DIVISION BY $J = 2^N$

(c) DIVISION BY FACTOR: (integer part J/K) +1

(d) DIVISION BY FACTOR $2^N/K$

FIG. 3L. TYPES OF BINARY RATE DIVIDER (BRD)

If J can be limited to a binary power, the circuit of Fig. 3L(b) can be used. Its simplicity depends on the fact that a counter which overflows effectively resets itself to zero.

For both the circuits just considered, a uniform input pulse train will result in a uniform output pulse train, while a jittery input train will emerge (reduced) with less jitter. This smoothing property of integer division is discussed later.

The circuit of Fig. 3L(c) is a modification of that in Fig. 3L(a) to allow introduction of an extra factor K. The operating rules are:

$$\begin{aligned} &(K + A) \rightarrow A \text{ ... on arrival of input pulse} \\ &(0 \rightarrow A \text{ and output pulse}) \text{., if } A \geqslant J \end{aligned} \qquad (3\text{-}37)$$

The division factor J is replaced [319] by: (smallest integer $\geqslant J/K$).

If the value of J is restricted to the binary power $M = 2^N$, so that the counter overflow property can be used in lieu of counter reset, the circuit of Fig. 3L(d) obtains. For this circuit, the operating algorithm is:

$$\begin{aligned} &(K + A) \rightarrow A \text{ ....on arrival of an input pulse} \\ &(A - M) \rightarrow A \text{ and an output pulse is generated ...} \\ &\qquad\text{....when } A \geqslant M. \end{aligned} \qquad (3\text{-}38)$$

A comparison of the algorithms (3-37) and (3-38) reveals a significant difference. With the latter the accumulator is not always zero following generation of an output pulse, but has a residual value between 0 and K-1. The size of this residue may affect the number of further input pulses before the next accumulator overflow is reached. The division ratio (i.e. input pulses per output pulse) may vary by 1. This ratio takes the two values, P_L and P_S, where

$$P_L = (\text{smallest integer} \geqslant M/K)$$

$$P_S = (\text{largest integer} \leqslant M/K).$$

This last circuit (Fig. 3L(d)) is the basic operational unit of many direct digital analysers (DDA's), where J is regarded as a variable, input pulses as increment Δx, output pulses as increments Δz, where $\Delta z = J \cdot \Delta x$. For DDA applications J would be so coded as to be able to represent both negative and positive numbers [304]. For pulse-frequency applications the problem of sign can usually be solved in other ways; the circuit is only required for division by the positive number M/K, which can also be thought of as multiplication by the fraction $K/M = K'$. Although this circuit is derived from a divider, and is hence called a 'binary rate divider', (BRD), it can be used as a direct substitute for the binary rate multiplier (BRM) described in the last section.

The multiplication factor, $K' = K/M$, of a BRD may have values from 0 to $1 - M^{-1}$ in steps of M^{-1}. The BRD thus avoids the unwelcome connection between factor and resolution that was discussed at the beginning of this section in connection with a simple divider.

As an example of BRD action, consider multiplication by the factor 3/8. That is, K = 3, M = 8 and therefore N = 3.
With the accumulator initially empty, i.e. $A_o = 0$:

Input pulse number	1	2	3	4	5	6	7	8
Accumulator content, A = 0	3	6	9/1	4	7	10/2	5	8/0
Output pulse ?			✓			✓		✓

With the accumulator not initially empty, e.g. $A_o = 5$:

Input pulse number	1	2	3	4	5	6	7	8
Accumulator content, A = 5	8/0	3	6	9/1	4	7	10/2	5
Output pulse	✓			✓			✓	

So changing the initial value in the accumulator alters the position but not the number of output pulses.

(i) Variance of BRD output spacing.

For M input pulses there will be K output pulses, and K interpulse spacings. Of these spacings, L will be oversize, equal to P_L / f, and S will be undersize, equal to P_S / f. The mean spacing is $\bar{\tau} = M / K \cdot f$. Writing M / K as $P + p$, where P is integer, and $0 \leqslant p < 1$, then

$$\left.\begin{aligned} P_S &= P \quad \text{and} \quad P_L = P + 1 \\ M / f &= L \cdot (P+1) / f + S \cdot P/f \\ K &= L + S \end{aligned}\right) \tag{3-39}$$

and solving these gives

$$L = M - K \cdot P \quad \text{and} \quad S = (P + 1)K - M. \tag{3-40}$$

The normalised variance (as eq. (3-26)) of a BRD output is thus:

$$\sigma_{\tau}^{2} = p\,(1-p) / (P+p)^{2} \tag{3-41}$$

which is plotted as the lower curve of Fig. 3I.

(ii) Patterning and low-frequency content of BRD output.

The output of a BRD cannot be simply reduced to its spectral components (as could the BRM output). However certain general properties are observable.

A BRD output pulse can only occur in a 'synchronous' time slot (i.e. coincident with an input pulse). The output pulse-train can thus be regarded as a slowly repetitive waveform (period = $M / f = 1 / f_1$) sampled by a high frequency pulse train (frequency = f). By sampling theory the

BRD output spectrum will therefore 'fold' about the input frequency, f : the power at frequencies 'a' and '(f-a)' will be the same.

A similar sort of folding relates to the scaling factor K / M. For K = X, X output pulses are generated in some pattern over $1/f_1$ seconds. For K = M-X, the output pattern shows X gaps in an otherwise uniform train of M pulses. The pattern of pulses for K = X and of gaps for K = M-X are the same, so that for all frequencies other than $j \times f_1$ where j = 0, 1, 2, ..etc., the output spectra are the same for K = X and K = M-X.

This folding with respect to frequency and factor is illustrated by the table below, wherein the modulus of the spectral components of a BRD output are plotted for all possible factors K. The table is for a 3-bit BRD, and to simplify the figures, Wf_1 = pulse area x input frequency / M, has been chosen to have value $\frac{1}{2}$ (amplitude unit).

Table : entries are modulus of Fourier components.

	K=0	1	2	3	4	5	6	7	8
frequency: d.c.	-	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
f_1	-	1.0	-	0.42	-	0.42	-	1.0	-
$2 . f_1$	-	1.0	2.0	2.0	-	2.0	2.0	1.0	-
$3 . f_1$	-	1.0	-	2.82	-	2.82	-	1.0	-
$4 . f_1$	-	1.0	2.0	2.0	4.0	2.0	2.0	1.0	-
$5 . f_1$	-	1.0	-	2.82	-	2.82	-	1.0	-
$6 . f_1$	-	1.0	2.0	2.0	-	2.0	2.0	1.0	-
$7 . f_1$	-	1.0	-	0.42	-	0.42	-	1.0	-
$8 . f_1 = f$	-	1.0	2.0	3.0	4.0	5.0	6.0	7.0	8.0

For an ideal rate multiplier a table of Fourier coefficients like that above would have dominant diagonal terms. If $Wf_1 = \frac{1}{2}$, as just considered, then denoting the modulus of the amplitude of the i^{th} harmonic component by a_i, for the ideal scaler:

$$\left.\begin{array}{lll} a_o & = K/2 & \\ a_{jK} & = K \quad \ldots\ldots & \text{where } j \text{ is integer} \\ a_{jK} & = 0 \quad \ldots\ldots & \text{where } j \text{ is not integer} \end{array}\right\} \text{— See sketch below.}$$

In particular, the amplitude modulus of the mean output frequency should satisfy:

$$a_K/K = 1.$$

For the 3-stage scaler of the table above:

K =	1	2	3	4	5	6	7	8
a_K/K =	1.0	1.0	.94	1.0	.56	.33	.14	1.0

For a 5-stage scaler of the same type (BRD):

K =	1	2	3	4	5	6	7	8	9	10	...
a_K/K =	1.0	1.0	.99	1.0	.96	.95	.92	1.0	.89	.85	...

In general, as $M \to \infty$, so

for K a factor of M:

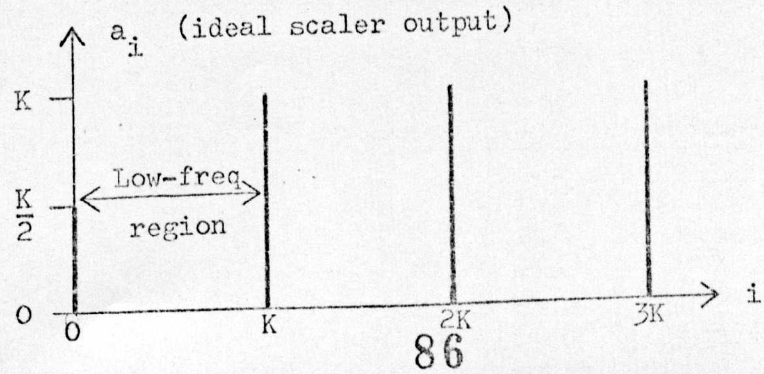
$$a_K = K, \quad \text{giving} \quad a_K/K = 1;$$

for K not a factor of M:

$$a_K = Wf \cdot \frac{2}{\pi} \cdot \text{Sin } (K'\pi)$$

giving here

$$a_K/K = \frac{\sin (K'\pi)}{K'\pi} \quad . \qquad (3\text{-}42)$$



The values of a_i in the low-frequency region ($i < K$) would be zero for an ideal scaler. With a BRD, a_i is a complex function of i, K and scaler capacity M. The analysis that follows is approximate, but substantially valid in the low-frequency region, where it yields upper bounds for a_i.

Using an ideal scaler, the j^{th} output pulse would occur at time: $t = j \cdot M / Kf = j / Kf_1$.

Using a BRD there will be pulse jitter: e, so that the j^{th} output pulse occurs at:

$$t_j = j / Kf_1 + e_j / f \qquad (3\text{-}43)$$

The displacement, e_j, is a linear function of the overflow left in the accumulator after the j^{th} pulse is output, this overflow being an integer in the range: 0 to K-1.

The overflows following successive output pulses can be shown to constitute a set of samples at times: $t = j / Kf_1$ of a sawtooth function of time, with amplitude: K-1 and period: $1 / Rf_1$, where

R is the Residue (M modulo K) (3-44)

Consequently the set e_j (where j = 1, 2, 3, ...) may be expressed as samples at times: $t = j / Kf_1$ of a sawtooth signal, e(t), of amplitude: $\pm\frac{1}{2}$ and also of period: $1 / Rf_1$.

Assuming correct phasing, e(t) can be expanded into the Fourier series:

$$e(t) = \frac{1}{2}\left(\frac{\sin[2\pi Rf_1 t]}{\pi/2} - \frac{\sin[4\pi Rf_1 t]}{2\pi/2} + \ldots \right) \qquad (3\text{-}45)$$

The amplitude of the m^{th} harmonic in the output $C_K(t)$, (factor = K/M) is:

$$a_{m,K} = 2f_1 \int_{1}^{1/f_1} C_K(t) \cdot \cos(2\pi m f_1 t)\, dt \qquad (3\text{-}46)$$

where the time origin is such that $C_K(t)$ is symmetric.

As

$$C_K(t) = W \sum_{j=0}^{K-1} \text{impulse } (t - \bar{t}_j), \quad \text{where } W \text{ is pulse area,}$$

equation (3-46) becomes

$$a_{m,K} = 2f_1\, W \sum_{j=0}^{K-1} \cos\left(2\pi \frac{m}{K} j + 2\pi \frac{m}{M} e_j\right)$$

$$\doteqdot 2f_1\, W \sum_{j=0}^{K-1} \left(\cos\left(2\pi \frac{m}{K} j\right) - 2\pi \frac{m}{M} \cdot e_j \sin\left(2\pi \frac{m}{K} j\right)\right)$$

provided $m \ll M$.

The first term of the expansion is zero, so

$$a_{m,K} = -4\pi f_1\, W \cdot \frac{m}{M} \sum_{j=0}^{K-1} \left(e_j \sin\left(2\pi \frac{m}{K} j\right)\right) \qquad (3\text{-}47)$$

The sum of sinusoidally weighted samples of a sinusoid, over an integer number of cycles of both, is only non-zero if both sinusoids have the same frequency. Allowing for aliasing when $e(t)$ is sampled at rate Kf_1 then from (3-45)

$$e_j = \frac{1}{\pi}\left(\sin(2\pi R f_1 \cdot j/Kf_1) - \sin(2\pi (K-R) f_1 \cdot j/Kf_1)\right)$$

+ smaller terms at other frequencies.

$$= \frac{1}{\pi} \sin\left(2\pi \frac{R}{K} \cdot j\right) - \frac{1}{\pi} \sin\left(2\pi \frac{K-R}{K} \cdot j\right)$$

$$- \frac{1}{2\pi} \sin\left(2\pi \cdot \frac{2R}{K} \cdot j\right) + \frac{1}{2\pi} \sin\left(2\pi \cdot \frac{(K-2R)}{K} \cdot j\right) \ldots \text{etc.}$$

$a_{m,K}$ will have maximum modulus when the fundamental or folded fundamental of e_j coincides with $\sin\left(2\pi\frac{m}{K}\cdot j\right)$, which is when $m = R$ or $m = K - R$.

Under these circumstances:

$$a_{m,K} = \pm 4\pi \cdot f_1 W \cdot \frac{m}{M} \cdot \frac{1}{\pi} \sum_{j=0}^{K-1} \sin^2\left(2\pi\frac{m}{K}\cdot j\right)$$

$$= \pm 2 f_1 W \cdot \frac{mK}{M} \quad \Bigg| \begin{array}{l} m = R \\ \text{or } m = K - R \end{array} \tag{3-48}$$

More generally, for K prime with respect to m, and neglecting higher order terms:

$$a_{m,K} = 2f_1 W \cdot \frac{mK}{qM} \quad \Bigg| \begin{array}{l} m = qR \\ \text{or } m = K - qR. \end{array} \tag{3-49}$$

The spectrum of a BRD output shown in Fig. 3J agrees with these formulae.

(iii) Pulse position errors of a BRD.

For a BRD where M/K is integer, the output pulses are evenly spaced. In general, however, M/K is not integer and the residue A_r in the accumulator, just following an output pulse, varies from pulse to pulse. If M is prime with respect to K, for M input pulses A_r will take once each the values 0 to K-1. As discussed earlier, the pulse position error as a fraction of an input period is A_r/K provided $A_o = 0$.

If the initial value in the accumulator is set to

$$A_o = (M + K - 1)/2 \tag{3-50}$$

the mean pulse position error will be zero, and the actual position errors will range uniformly over the range

$$-1/2f < \text{pulse position error} < 1/2f \tag{3-51}$$

or more exactly

$$-(K-1)/2Kf \leqslant \text{error} \leqslant (K-1)/2Kf \tag{3-52}$$

For this quantised boxcar distribution, and normalising with respect to mean period $M/K.f$,

$$\sigma_x^{\,2} \doteqdot \frac{(K-1)^2}{12\,K^2 f^2} \Bigg/ \frac{M^2}{K^2 f^2} \doteqdot \frac{(K-1)^2}{12\,.\,M^2} \,. \tag{3-53}$$

This normalised pulse position error variance is maximum for $K = M-1$ when:

$$\sigma_x^{\,2}\Big|_{max} \doteqdot 1/12 \doteqdot 0.083 \,. \tag{3-54}$$

Mean variance over all factors

$$\sigma_x^{\,2}\Big|_{mean} \doteqdot \frac{1}{M}\int_0^M \frac{(K-1)^2}{12M^2}\,.\,dK \doteqdot \frac{1}{36} \tag{3-55}$$

$$= 0.028$$

Equations (3-54) and 3-55) give the asymptotes to the computed results plotted in Fig. 3K.

(iv) Staircase error of a BRD.

From the discussion of the last two subsections, the mean staircase error is zero ($\bar{y} = 0$), provided the correct initial conditions are set, viz. $A_o = (M+K-1)/2$. Under these conditions the maximum staircase error is

$$y_m = K-1/2K + 1/2 \doteqdot 1 \quad \text{for } K = M-1 \tag{3-56}$$

3.4.4 Comparison of BRM and BRD.

By all the criteria proposed in section 3.4.1, the BRD is superior to the BRM as a frequency scaler. Taking the criteria in turn:

(i) Variance of period, equations (3-26) and (3-41), and Fig. 3I.

For scale factors K' $(= K/M) > \frac{1}{2}$, both scalers give the same period variance that rises to 0.125 at $K' = 0.75$. For smaller scale factors $(< \frac{1}{2})$, the BRD period variance falls off sharply with the factor, so that $\sigma_\tau^{\,2} < (K')^2/4$. Only for very small factors (K') could the factor be reliably deduced from a single output period measurement.

(ii) Low frequency spectral content, equations (3-30) and (3-49), and Fig. 3J. The BRM spectrum is very different from that of an ideal scaler, for not only is there often no significant concentration of power at the mean output frequency, Kf_1 (or K'f), but also there is considerable power at frequencies below Kf_1. These 'subharmonic' spectral components can be given an upper bound as follows:

$$a_{m,K}\Big|_{max} = (2\,Wf_1 \times m), \quad \text{for BRM}$$

$$a_{m,K}\Big|_{max} = (2\,Wf_1 \times m \times \frac{K}{M}), \text{for BRD} \qquad (3\text{-}57)$$

provided $m < K$; $K < M/2$.

For both types of divider, $a_{m,K} = a_{m,\overline{M-K}}$.

For both types of divider, the spectrum is the same if M/K or $M/(M-K)$ is an integer; but otherwise the amplitude of BRD subharmonic components average less then half the amplitude of corresponding BRM components. A dynamic filter (e.g. phase locked loop) would be able to smooth a BRD output much more easily than a BRM output.

(iii) and (iv) Pulse position and staircase errors, equations (3-32), (3-34), (3-35) and (3-54), (3-55), (3-56) in conjunction with (3-22); see also Fig. 3K. The errors for a BRD tend to values that are not functions of scaler length: N. BRM errors, by contrast, are increasing functions of N. The table over summarises these errors.

Table of pulse-position and staircase errors for BRM and BRD

	error		error ratio BRM : BRD		
	BRM	BRD	N=5	N=10	N=15
Pulse position error: x		(asymptote)			
max (given $\bar{x}=0$): x_m	$0.167N - 0.11$	0.5	1.5	3.1	4.8
max variance (σ_x^2)	$0.028N - 0.019$	0.083	1.5	3.1	4.8
mean variance "	$0.020N - 0.034$	0.027	2.4	6.0	9.6
Staircase (integral) error: y					
max (given $\bar{y}=0$): y_m	$0.167N + 0.39$	1.0	1.3	2.1	2.9
max variance (σ_y^2)	$0.028N + 0.064$	0.167	1.3	2.1	2.9
mean variance "	$0.020N + 0.049$	0.111	1.4	2.2	3.1

Examination of the table suggests the empirical rule

$$x_m = 6 \cdot \sigma_x^2 \Big|_{max} \quad \text{or} \quad y_m = 6 \cdot \sigma_y^2 \Big|_{max} \qquad (3\text{-}58)$$

which has not however been derived from the theory. Such a rule does however agree with equation (3-22) linking x and y.

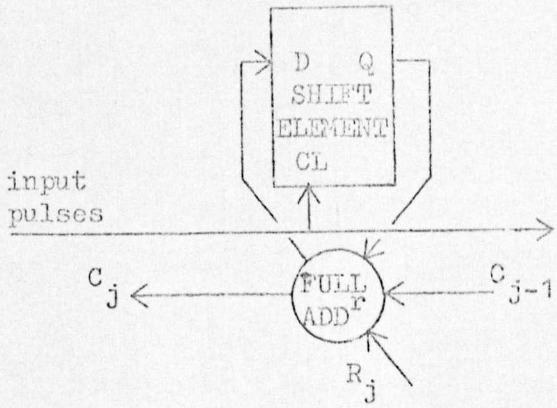
A practical choice between a BRM and a BRD may depend upon criteria other than the error criteria above. The components to construct one stage of each type of scaler are:

BRM (see Fig. 3H(a)): 1 bistable; 1 differentiating circuit; 1 AND gate

BRM (see Fig. 3H(b)): 1 bistable; 1 shift register; 1 AND gate

BRD (see Fig. 3L(d)): 1 shift register element; 1 full adder

The BRD circuit is shown more explicitly here; the output is the carry C_N from the final stage, (gated by the input pulses or shaped by a monostable).



Thus the component cost of a dedicated BRM and BRD scaler will be similar. However where a single frequency is to be multiplied in parallel by a number of constants K_1 to K_x a BRM can be divided so that a single set of bistables and differentiating circuits feed X sets of AND gates. The BRD circuit cannot be shared in this way.

A particular design of a divided BRM has been used in a multi-port digital-to-analogue convertor [321]. X (N-bit) factors K_1 to K_x are stored in X shift registers, each initially with its most significant bit at the output of the shift register. The shift registers are individually cycled in an ingenious way upon arrival of each of a train of clock pulses, so that at the output of (for example) the J^{th} register, the following sequence is observed (suffices denote the bits of the binary number K_J)

Sequence number:	1	2	3	4	5	6	7	8	....
Bit presented to output:	k_N	k_{N-1}	k_N	k_{N-2}	k_N	k_{N-1}	k_N	k_{N-3}	

The sequence repeats after $M = 2^N$ clock pulses. At each clock pulse, an output pulse is generated on line J if the bit present at the output of shift register J is a 1. It may be ascertained that the mean output frequency on line J is clock frequency x K_J / M, and similarly for all X lines. The output pulse pattern is the same as would be obtained from a set of X BRM's each fed with the same clock train. The complexity of the cycling arrangements for the shift registers is accepted because it allows each channel to consist only of one shift register and one AND gate to perform both digital storage and digital-to-frequency conversion. Frequency-to-analogue conversion requires only a low pass filter. To attempt this technique with a BRD method would involve about twice as many components.

A further minor advantage of a BRM is that for zero mean staircase or pulse position error, the initial state should be zero ($A_o = 0$), which is easy to arrange. For a BRD the initial state should be $A_o = (M+K-1)/2$, which is less easy to arrange. In continuous applications, where K varies with time, 'initial state' has no meaning.

The construction of a BRD allows the addition of 1 to the scaling constant K, so that the effective scaling factor is $(K+1)/M$. There are a number of situations where this facility is useful. The same facility can be added to a BRM by incorporating an extra stage.

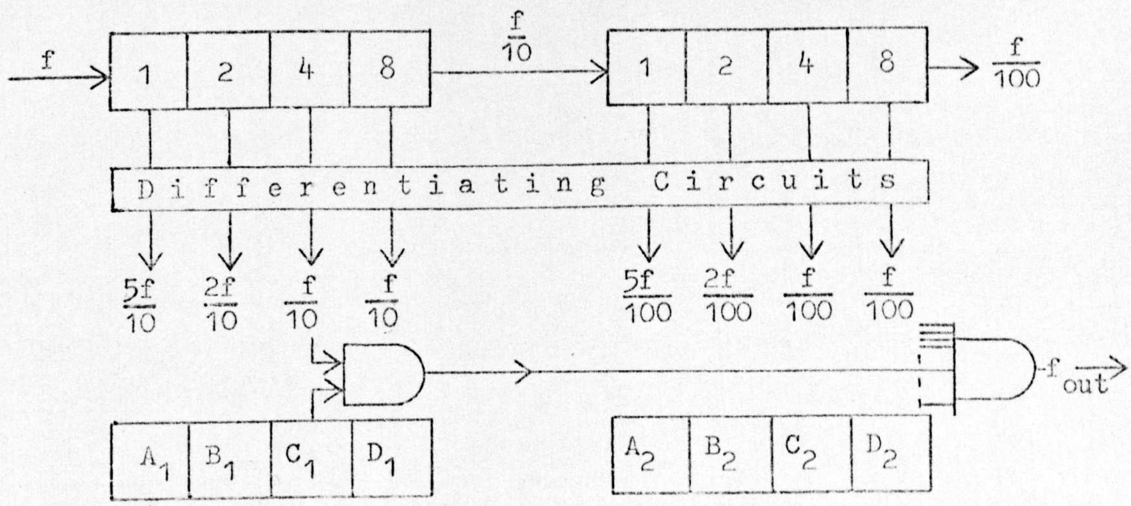
3.4.5 Other digital scalers.

A number of modifications to the basic BRM have been proposed, while binary coded decimal rate multipliers are also used.

Talbot and Senior [323] proposed a BRM with negative feed back to obtain an overall scaling factor of $1/(1+K')$ where K' = (presented binary integer)/(capacity of principal register) = K/M. They showed the staircase error of this composite scaler never exceeds 1 step. Such a scaler is not very useful, however, as the relation between effective factor and K is complex, and the effective factor is limited to the range $\frac{1}{2}$ to 1.

Positive feedback has been variously suggested for obtaining scale factors of greater than unity, i.e. of $K'/(1-K')$. As output pulses are fed back to the scaler input, they must be passed through a fixed delay of T_D. Consequently, regardless of input frequency, some of the output pulses will be separated by T_D. The spectral purity (etc.) of the output train will be very poor. Again the relationship between effective scaling factor and K is not a convenient one.

Decimal rate multipliers [322, 324, 325] use binary coded decade stages for their divider portion (c.f. the pure binary stages of a BRM). The <u>transitions</u> of the bits of a binary counter, counting at constant rate, can be used to generate the binary subharmonics of that rate. The transitions of the bits of a binary coded decimal counter do not generate binary subharmonics. Consider for example an 8421 coded counter that resets after reaching state 9 '1001'. For every 10 input pulses, the 8, 4, 2, and 1 counter bits make 1, 1, 2 and 5 'α' transitions respectively, and also generates 1 carry pulse. None of these transitions coincide, so they can be used to trigger pulses into four gates, in the same manner as a BRM.



In the diagram above,

$$f_{out} = f \left[A_1 \cdot \frac{5}{10} + B_1 \cdot \frac{2}{10} + C_1 \cdot \frac{1}{10} + D_1 \cdot \frac{1}{10} \right.$$

$$\left. + A_2 \cdot \frac{5}{100} + B_2 \cdot \frac{2}{100} + C_2 \cdot \frac{1}{100} + D_2 \cdot \frac{1}{100} \right]$$

$$f_{out} = f \left[\frac{(5A_1 + 2B_1 + C_1 + D_1)}{10} + \frac{(5A_2 + 2B_2 + C_2 + D_2)}{100} \right] \quad (3\text{-}39)$$

The decimal rate multiplier thus requires the coded digital scaling factor to be coded differently to the divider that generates the sub-harmonic pulse trains. Kostka [325], amongst others, has listed 'complementary' code pairs suitable for use in BCD rate multipliers.

Divider code				Scale factor code			
1	2	4	8	5	2	1	1
1	2	2	4	5	2	1	1
1	2	3	3	4	3	1	1
1	1	3	4	3	3	2	1
1	1	2	5	4	2	2	1

Certain codes, (e.g. certain forms of 1224 code) are unacceptable for the divider because more than one bit may make an $\propto$ transition at the same time.

Kaps [324] examined different types of BCD rate multipliers and concluded that they all yield substantial (staircase) errors, particularly when the scaler is several decades long. He proposed a modification to the divider of Fig. 3L(a), whereby a feedback signal is used to modify the divisor J. The principle of his modification can be seen from the following example:

Suppose the desired division factor is 2.4. This may be realised if J (which is an integer division factor) is varied between 2 and 3 in a systematic way. For example if for $\frac{6}{10}$ of the output pulses the division factor is 2, and for $\frac{4}{10}$ of the output pulses the division factor is 3, then per 10 output pulses there would have been 24 input pulses (i.e. 6x2 + 4x3). The output pulse train will be more uniform if the changes of division factor J (between 2 and 3) are made as frequently as possible, e.g.

Input pulse	1	2	3	4	5	6	7	8	9	10	11	12	
J	2 ———				3 ———			2 ——		3 ———			etc.
Output pulse		1		2			3		4			5	

The technique is not simple, but does allow division by a non-integer, and does yield a comparatively even output pulse train. The technique is only indirectly applicable to rate multiplication: a Kaps type divider by factor M/K would give a more even output than a BRM with factor K/M.

3.4.6 Smoothing by division.

If a pulse-train with jitter is passed through a divider (i.e. every N^{th} pulse is selected), the resultant pulse train will naturally have a lower mean frequency ... it will also exhibit relatively less jitter. Thus the output from a scaler can be smoothed by division. Where the system allows, therefore, it is worthwhile to operate scalers at higher frequencies than is finally required, and to pass their outputs through (integer) rate dividers. Scaling followed by fixed-factor division is preferable to the division followed by the scaling.

The integer division-factors most simply realised are the members of the binary power series: 2, 4, 8, 16, etc., and these will be considered in greater detail.

The output of a BRM or BRD that is fed with a uniform pulse train will be patterned, the pattern repeating after every P pulses, where P is odd. (If the scaling factor is K/M, P is the only odd integer in the series K, K/2, K/4, K/8, etc.). Division of such a scaler output by 2 (and reshaping the pulses, else a distorted square wave results), results in a doubling of the mean period and changes in the pulse pattern. However the number of pulses in a complete pattern is the same as before, viz. P.

The pulses of the original pattern were displaced about their true positions by times: $e_j \cdot \bar{\tau}$, $(j=0,1,2,..,P-1)$, where $\bar{\tau}$ is the mean period. After division, the displacements will be the same set $e_j \cdot \bar{\tau}$, although probably in a different order. As the mean period after division is $2\bar{\tau}$, the normalised values of the pulse-position error will be halved by division (and halved by each further such division). In particular the normalised maximum pulse-position error will be halved and the normalised pulse-position-variance quartered by each division. The staircase errors (y) will follow the normalised pulse position errors (x) according to equation (3-22).

The effect of division upon period and low frequency spectral content is beneficial, but less easy to quantify than the effect on pulse position errors.

For a BRM, the output period takes two values in a ratio of 2 : 1 that straddle the mean output period $\bar{\tau}$. After division by 2,4,8, etc. either two or three values of period are observed in the ratios of two or three adjacent integers (e.g. 2 : 3 : 4), whose mean values approximately double with each division.

For a BRD, the output period takes two values in the ratio of two adjacent integers. After each division, two new such values are observed, the integers again approximately doubling.

In both cases (BRM and BRD), the actual output periods bracket the desired output period more closely with each division. Fig. 3M shows this effect in its lower two rows.

If 'z' is the error in the period of a random, but statistically stationary, pulse train, then the normalised variance of that period can be denoted:

$$\sigma_\tau^2 = \langle z^2 \rangle / \bar{\tau}^2, \quad \text{where } \bar{\tau} \text{ is the mean period.}$$

Following division of the pulse train by 2, and using the notation 'z_{1+2}' to denote the sum of the errors in two successive periods, then the normalised variance of period, τ is:

$$\sigma_{2\tau}^2 = \langle (z_{1+2})^2 \rangle / 4\bar{\tau}^2 = \tfrac{1}{2}\sigma_\tau^2 \left(1 + \phi_{zz}(1)/\phi_{zz}(0)\right) \qquad (3\text{-}60)$$

The function: $\phi_{zz}(a)$ is the expected value of the cross-product of the errors in output periods separated by 'a' pulses. Thus, $\phi_{zz}(0) = \langle z^2 \rangle$. The value of this auto-correlation function for the deterministic output of a scaler with factor, $K' = K/M$, is:

$$\phi_{zz}(a) = \frac{1}{K} \sum_{j=1}^{K} (z_j . z_{j+a}) \quad . \qquad (3\text{-}61)$$

So defining 'R':

$$R = \phi_{zz}(1)/\phi_{zz}(0)$$

then equation (3-60) can be rewritten:

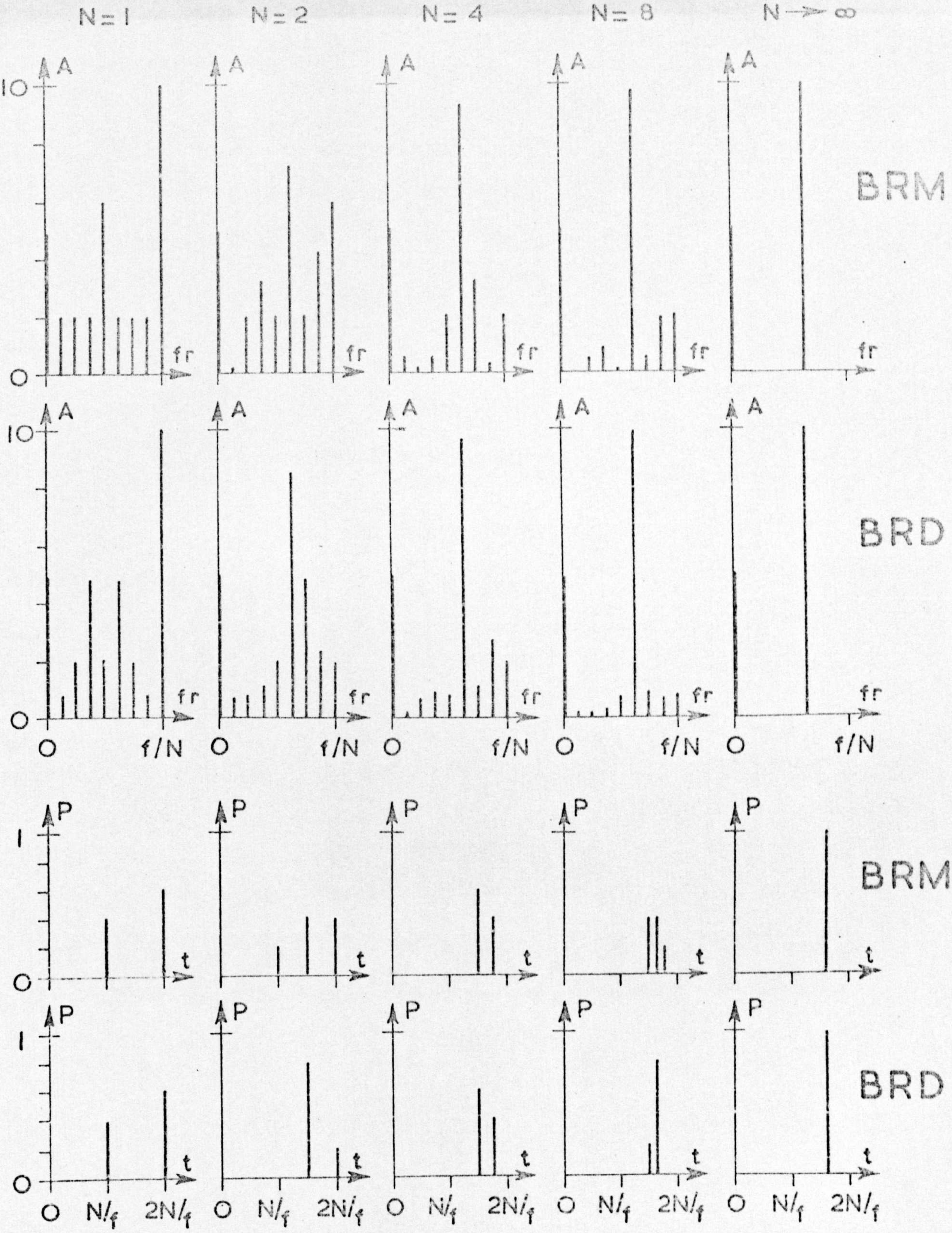
$$\sigma_{2\tau}^2 / \sigma_\tau^2 = \tfrac{1}{2}(1+R) \quad . \qquad (3\text{-}62)$$

For a train of Poisson-distributed pulses, R = 0; so division by 2 reduces the normalised variance of period by 2. For a BRM output, R takes values in the range: $-\frac{1}{2}$ to $+\frac{1}{2}$; so division of the pulse-train by 2 gives a variance reduction factor between 4/3 and 4. A BRD output yields R value between 0 and -1 , according to multiplication factor, K/M; the variance reduction factor consequently ranges upwards from 2. Thus a single division may be expected to improve the uniformity of BRD output by more than it does that of a BRM output.

Division by large numbers has much the same effect on both outputs, and upon Poisson-distributed pulse trains, viz:

$$\sigma_{N\tau}^2 \to \sigma_\tau^2 / N^2, \quad \text{when} \quad N \gg 1 . \tag{3-63}$$

The spectrum of divided scaler outputs has not been calculated. The top rows of Fig. 3M show that each successive division of a BRM or BRD output train tends to concentrate spectral power at the mean output frequency, and reduce subharmonic power.



A IS AMPLITUDE OF FOURIER COMPONENT (PLOTTED V. FREQUENCY)

P IS PROBABILITY OF OCCURRENCE OF A PERIOD (PLOTTED V. TIME)

ORIGINAL SCALER: INPUT FREQUENCY = f, FACTOR = 5/8

FIG. 3M. EFFECT OF DIVIDING SCALER OUTPUT BY N.

3.5 Pulse-cancelling and rate-comparison circuits.

The subtraction of one pulse-train from another calls for some form of pulse-cancelling or pulse-pairing circuit. Analogue pulse-cancelling circuits are complex and unreliable, as was discussed in section 3.2.2. A digital method of cancellation (wherein each pulse is treated as an event) is required.

Consider two pulse signals a(t) and b(t), whose pulses do not overlap. Their mean pulse rates may be denoted f_a and f_b, while the number of pulses since the time origin (t=0) are A and B respectively. A pulse canceller is a circuit with two output signals c(t) and d(t), whereby

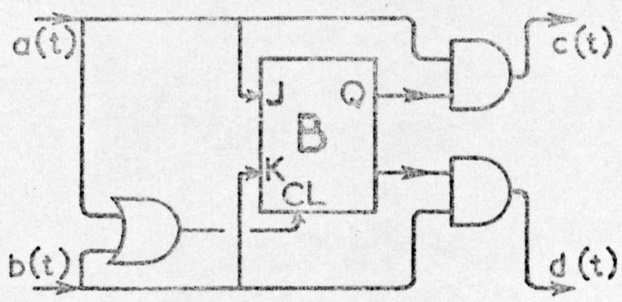
$$f_c = f_a - f_b\ ;\ C = A - B;\ f_d = 0; \qquad D = 0 \qquad \ldots \text{ if } f_b < f_a$$

$$f_c = 0; \qquad C = 0; \qquad f_d = f_b - f_a;\ D = B - A \ \ldots \text{ if } f_b > f_a \tag{3-63}$$

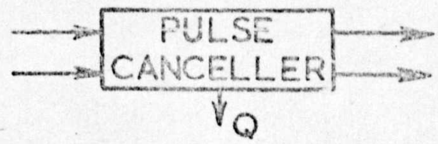
These relationships assume both a(t) and b(t) pulse trains are evenly spaced. Such a canceller is shown in Fig. 3N(a) and the corresponding waveforms in Fig. 3N(c). The output signals are pulse-trains with uneven pulse-spacing.

If the input trains a(t) and b(t) were irregularly spaced, there is a possibility of outputs on both c(t) and d(t) lines, whereupon equations (3-63) above do not hold. In such cases more than one pulse-canceller is necessary, and if a sufficient number are cascaded equations (3-63) will apply. If x_m is the maximum pulse-position error of a jittery pulse-train, then the number of pulse-cancellers to reliably subtract pulse-train b(t) from train a(t) is R where:

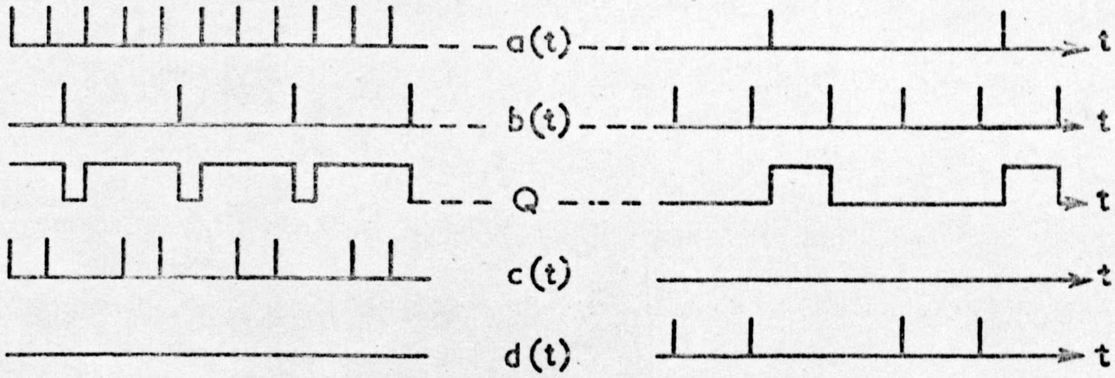
$$R = 2 + \text{integer part } (x_{m,a} + x_{m,b}) \tag{3-64}$$



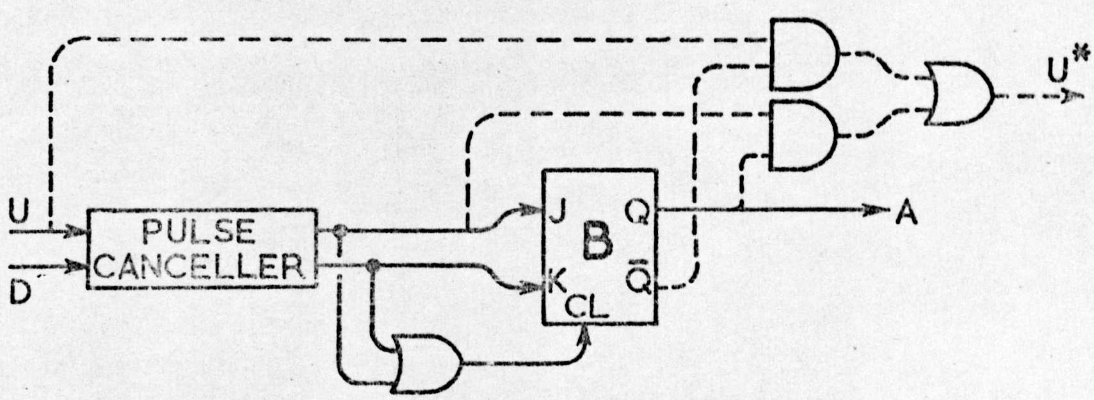
(a) PULSE CANCELLER



(b) GENERAL SYMBOL

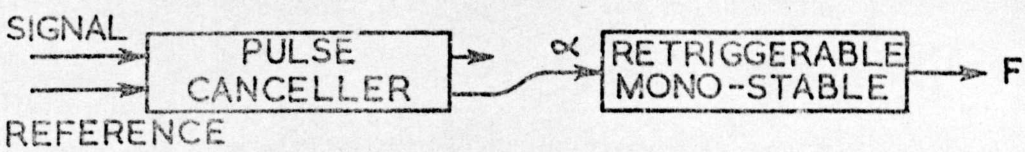


(c) WAVEFORMS OF INPUT AND OUTPUT PULSE-TRAINS.



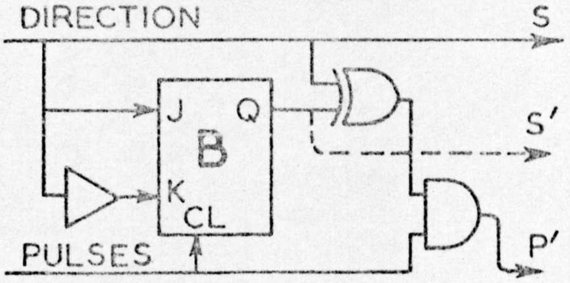
(d) FREQUENCY COMPARATOR $f_U > f_D$: $A = 1$ $(f_{U*} = f_U - f_D)$

$f_U < f_D$: $A = 0$ $(f_{U*} = f_U)$

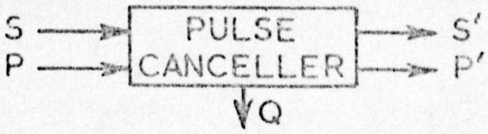


(e) LOWER FREQUENCY LIMIT DETECTOR.

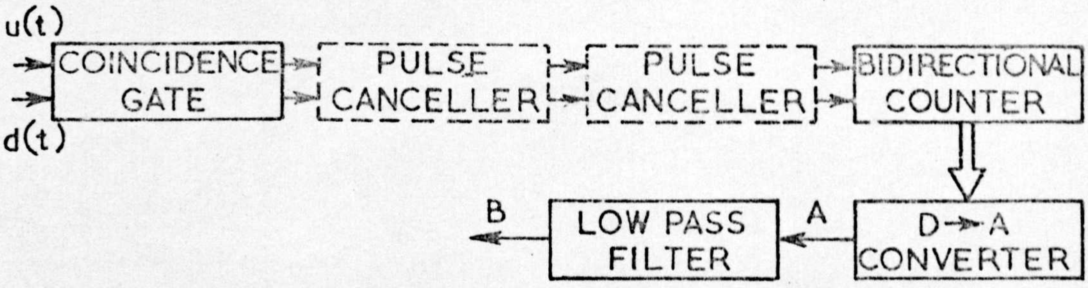
FIG. 3N. PULSE-CANCELLER CIRCUITS.



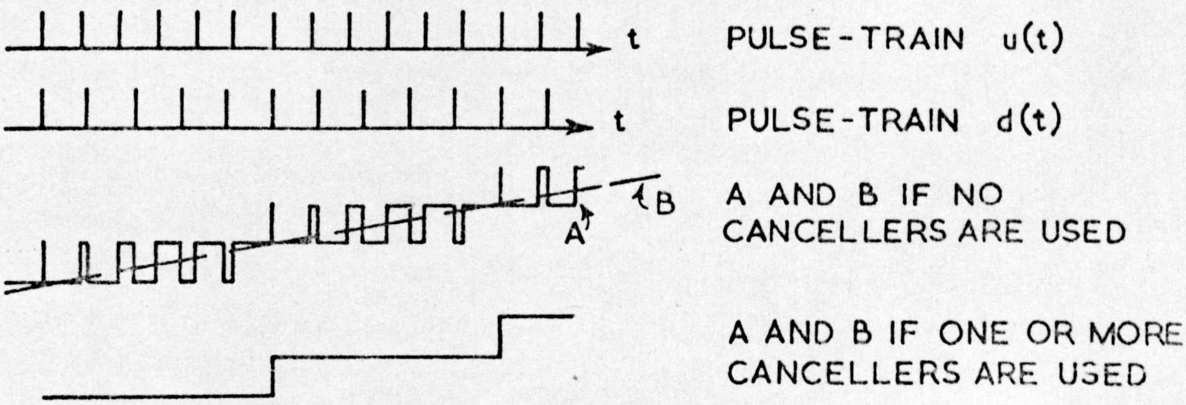
(a) PULSE-CANCELLER



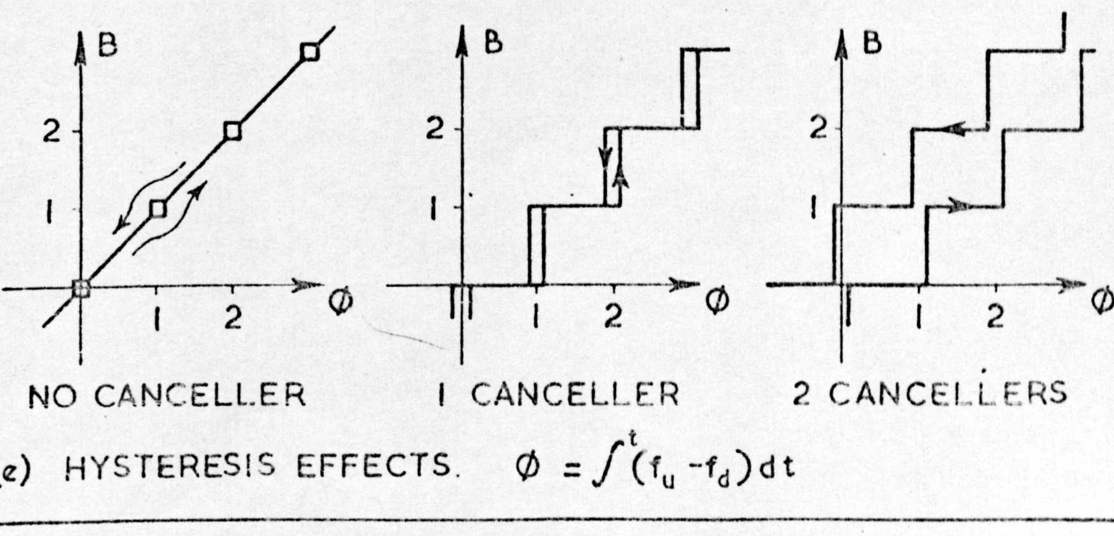
(b) GENERAL SYMBOL



(c) COUNTING THE DIFFERENCE BETWEEN TWO PULSE-TRAINS



(d) WAVEFORMS FOR (c) ABOVE



(e) HYSTERESIS EFFECTS. $\phi = \int^{t} (f_u - f_d)\, dt$

FIG. 3P. FURTHER PULSE-CANCELLER CIRCUITS.

The frequency comparator of Fig. 3N(d) also requires R pulse-cancellers to function correctly. Equation (3-64) quantifies the significance of the pulse-position errors in scaler outputs, which were examined at length in Section 3.4 above. This relation is hinted at by Rey[326], when he tabulates the excursions of a bidirectional counter that is fed from a scaler circuit.

The pulse-canceller illustrated in Fig. 3N(a) has been frequently described [326 - 331]. It is easily constructed using an integrated circuit 'master-slave' bistable, which will adequately resolve the incipient race problem. The pulse-canceller of Fig. 3P(a) fits a different sign convention; it is more easily appended to a coincidence--prevention gate. Sign conventions are discussed in Section 3.6 .

The arrangement of Fig. 3P(c) has been used in various control schemes [328 - 330]. It is introduced here to illustrate hysteresis effects caused firstly by coincidence-prevention gates, and secondly by pulse-cancellers. For the purposes of discussion, the two input pulse trains, $u(t)$ and $d(t)$, are assumed even and of unvarying frequencies f_u and f_d respectively. It is also assumed that $f_u > f_d$.

A relative 'phase', $\emptyset(t)$, may be defined as the time integral of the relative frequency, f_r . ($f_r = f_u - f_d$). As f_r is here constant, $\emptyset(t)$ increases linearly with time, and has a numerical value equal to the number of pulses the $d(t)$ train has slipped with respect to the $u(t)$ train.

If the incoming pulses have been conditioned by a coincidence gate not to approach within a small interval 'T' of each other, so as not to overlap, then $\emptyset(t)$ cannot take values within 'T' of any integer. The effect of the coincidence-prevention gate is to temporarily constrain the apparent value of relative frequency f_r to zero, as $\emptyset(t) \longrightarrow$ integer.

The apparent value of $\emptyset$ can be measured using a counter, a digital-to-analogue converter and a smoothing filter. This is illustrated in Fig. 3P(c) and (d). The resultant jumps in the apparent value of $\emptyset$, due to the coincidence of gate action, are plotted in the left-hand curve of Fig. 3P(e). The jumps are of size: 2Tf.

When a cancelling circuit is inserted, information about the gradual increase of $\emptyset$ is lost. Only when the faster pulse train has gained one complete pulse (period) on the slower train, will the counter increment or decrement. When two cancelling circuits are cascaded, a hysteresis of 1 count is introduced (to add to the 2Tf hysteresis already discussed). If the faster pulse train should become the slower, it must drop back two pulses (periods) before the counter changes. For K pulse cancellers, a phase hysteresis of K results.

When a frequency comparator (e.g. Fig. 3N(d)) is preceded by K pulse cancellers (to ensure steady output in the presence of pulse jitter), then a considerable delay in operation may ensue. If relative frequency f_r changes from positive to negative, the frequency comparator will take up to time $(K+1)/f_r$ to record the change.

The phase hysteresis effects just described are difficult to interpret in a pulse-frequency system, where they represent delays that are signal dependent. The phase jumps due to avoiding coincidence are usually insignificant, although one designer has sought to reduce them [329]. The phase jumps due to using many pulse-cancellers may lead to small amplitude oscillation in pulse-frequency control loops.

Pulse-cancellers form the basis of a number of pulse operational units. The comparator of Fig. 3N(d) can be extended in a number of ways.

The circuit shown dotted is a conditional frequency subtractor. If $f_u < f_d$, then the comparator output is '0', and f_u is passed on as f_{u*}. If $f_u > f_d$, the comparator output is '1', and $f_r = f_u - f_d$ is passed on as f_{u*}. This unit is used in a frequency tracking circuit described in the next section.

A cascaded series of pulse-cancellers may be regarded as a form of counter. The states Q of successive cancellers can be viewed as a binary count, with the Q of the first canceller as least significant bit. This counter has the property of counting in exponential steps which are reset to 1 following each change in count direction. For example applying 4 'up' pulses to an initially empty counter gives the sequence: 0001, 0011, 0111, 1111, while subsequently applying 4 'down' counts gives the sequence: 1110, 1100, 1000, 0000. These count steps are of sizes 1,2,4,8, -1,-2,-4,-8 respectively. By reversing the direction of count following each overshoot, the counter can be brought to a chosen state in relatively few steps. To get from any state to any other state with an 'n' bit version of such a counter requires not more than $n + (n-1) + (n-2) + \ldots + 1 = \frac{1}{2}n(n+1)$ steps. This compares favourably with the maximum of (2^n-1) steps for a conventional binary counter. Corresponding average numbers of steps between two arbitrarily chosen states are approximately $\frac{1}{4}n(n+1)$ and $\frac{1}{4}(2^n-1)$.

So we have a special counter that has the properties: (i) the count changes in the same direction as the incoming increment or decrement pulses (although not proportionally), (ii) on average a change in state requires fewer steps (as $\frac{1}{4}n(n+1)$: $\frac{1}{4}(2^n-1)$) than for a conventional bidirectional counter, (iii) only one bit changes at a time. However, (iv) to change a count by 1 may take several steps and involve substantial excursions.

As illustration of this last point, consider the change 15→16, which involves the sequence 01111, 11111, 11110, 11100, 11000, 10000; from 15 to 16 is via 31 ! This same phenomenon albeit only momentary has already been discussed with respect to ripple-through counters (section 3.3).

The frequency comparator of Fig. 3N(e) can be used as a rapid limit detector, responding within one reference train period to a fall in signal train frequency. It does not need to be preceded by a coincidence prevention gate. It differs from the more usual frequency comparator (Fig. 3N(d)) in that a momentary loss of signal results in output F remaining in the '1' state for some time. F does not oscillate if signal and reference frequencies are close to each other. If the signal pulse-period is consistently or even intermittently longer than the reference pulse-period, F remains at '1'. A further version of this circuit is obtained by replacing the mono-stable by a bistable. F then indicates any historic falling of the signal frequency below its lower limit (viz. the reference frequency), until it is externally reset to '0'.

3.6 The treatment of sign.

A variable represented by a pulse-frequency signal, (so that the mean frequency is proportional to the variable's value), cannot take a negative value: for negative frequency cannot be given direct physical meaning. If negative valued variables are likely to occur in a PFM system, some supplementation of the basic coding is required.

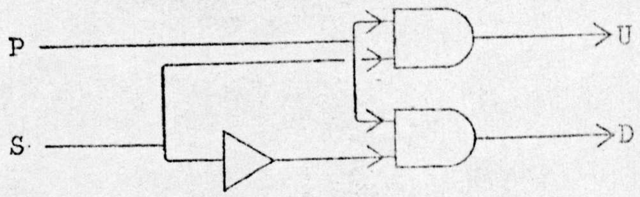
One technique is the use of bipolar pulses, the polarity of the pulses indicating the sign of the variable, while the pulse-rate indicates the modulus of the variable's amplitude. Such a technique essentially uses ternary devices, and considerable research (especially in France) has been directed at the problems of such devices. However the comparative simplicity and therefore reliability, cheapness and availability of binary logic devices call for binary solutions to the problem of sign. Three solutions will be considered, all of them in effect use two binary channels to carry one ternary signal.

Firstly a bias can be used. A positive bias applied to an analogue variable prior to encoding by PFM is equivalent to a positive bias frequency 'added' to the encoded signal's frequency. By such means the problem of sign is evaded, the bias being chosen larger than the modulus of the minimum expected value of the variable. In interpreting a coded signal, one must also have access to the bias as a frequency, which may therefore be considered as travelling on an adjacent line. The bias system is neither convenient nor complete. The inconvenience of adding and subtracting the bias is complicated if the PFM signal is at any time scaled; and the bias method alone will not handle the subtraction of PFM signals: (A + bias) - (B + bias) may be negative even where both (A + bias) and (B + bias) are positive. The 'housekeeping' in a system involving several variables and several arithmetic operations becomes most laborious if every variable is biased.

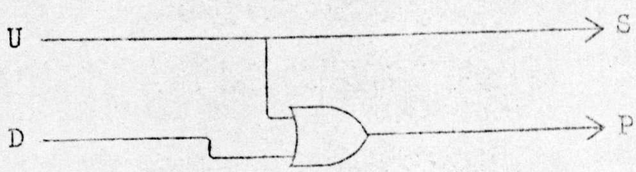
Secondly an auxiliary sign line can be introduced. The modulus of a variable's amplitude is represented by the pulse-frequency on a pulse (P) line, its sign by the logic state of a sign (S) line. Such a system may be denoted 'P&S'. For unambiguous operation, S should not change state during a P pulse, but rather between pulses.

Thirdly a balanced two line channel can be devised, the lines being denoted (say) U for up and D for down. If the variable is positive, pulses travel on the U line; if the variable is negative, on the D line. In either case the pulse-rate is proportional to signal modulus. It is usually advisable and easy to ensure pulses do not occur simultaneously on both lines. This mode will be denoted 'U&D'.

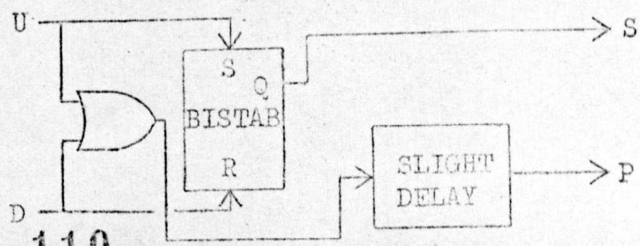
Conversion from 'P&S' to 'U&D' conventions is straightforward,



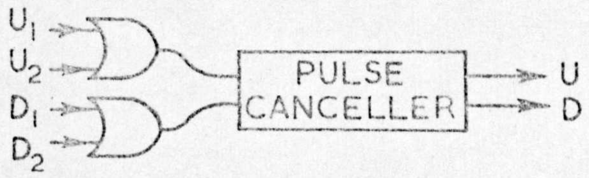
Conversion from 'U&D' to 'P&S' is a little more complex. The simplest circuit is open to the objection that



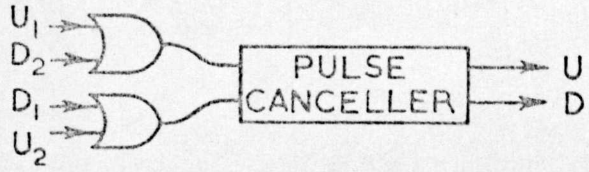
it gives a sign 'S' output which changes with the pulse 'P' output, and with a little system delay could even change during a pulse output. A safer circuit introduces a small deliberate delay so that there is no danger of S changing while P is high.

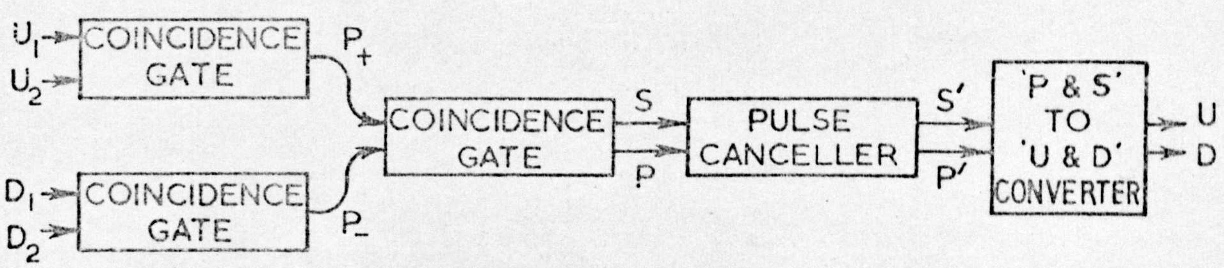


(a) ADDITION (U & D)

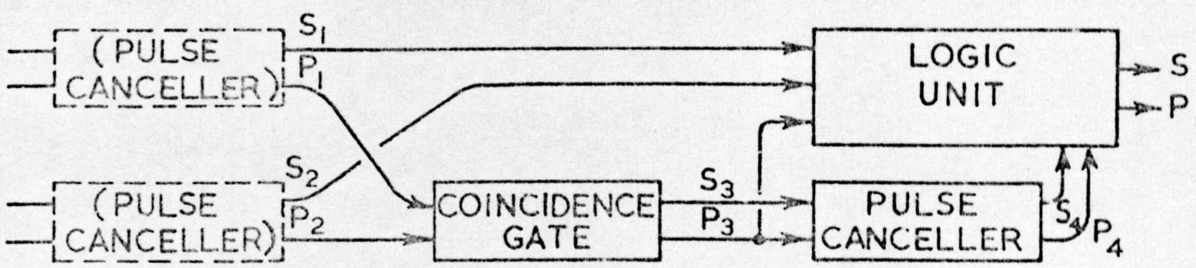


(b) SUBTRACTION (U & D)





(c) ADDITION (U & D) OF POSSIBLY COINCIDENT PULSE-TRAINS.



		$S_1\ S_2 \rightarrow$ 0 0	1 0	1 1	0 1
ADDITION	S =	0 (or S_1)	S_4	1 (or S_1)	$\bar{S}_4$
	P =	P_3	P_4	P_3	P_4
SUBTRACTION	S =	$\bar{S}_4$	1 (or S_1)	S_4	0 (or S_1)
	P =	P_4	P_3	P_4	P_3

(d) ADDITION & SUBTRACTION (P & S)

FIG. 3Q. ADDITION AND SUBTRACTION CIRCUITS.

This problem of conversion highlights a wider inconvenience of the 'P&S' notation. In passing through an anti-coincidence circuit or a synchronising circuit, pulses may be slightly delayed; the delayed P pulse may then overlap a transition on the S line. Thus, where possible, the S line transitions should not occur immediately after a P line pulse, but rather just before the following P line pulse. Alternating up and down pulses cause frequent transitions of the sign line, S, and where a 1 pulse hysteresis is acceptable, the insertion of a pulse canceller (to stabilise the S line) is advisable.

Addition and subtraction of non-overlapping pulses is quite straightforward with two 'U&D' signed pulse-trains. Suitable circuits are shown in Fig. 3Q(a) and (b). If however the pulse-trains might overlap, three coincidence gates described earlier give a 'P&S' signed output; a 'P&S' to 'U&D' converter is also shown in the diagram. Coincidence gates are fairly elaborate devices, so that to use three to perform one addition makes the operation a costly one.

Addition or subtraction with 'P&S' signed signals only requires one coincidence gate, but a number of other logical operations must also be implemented. The basic circuit, and a truth table for the logic unit are shown in Fig. 3Q(d).

Both 'P&S' and 'U&D' sign conventions occur naturally in a PFM computer, and may be interchanged fairly simply. Where overlapping pulses occur, or where bidirectional counters or scalers appear, the 'P&S' convention has advantages. In a synchronous system the 'U&D' convention is preferable. In any case operations upon signed variables should be avoided wherever possible.

3.7 Frequency-to-number conversion

The magnitude of a frequency is a continuous variable, and frequency has physical units T^{-1}. So frequency-to-number conversion entails the two actions: comparison with a physical standard and rounding to an integer.

The most convenient form of the conversion standard is a reference or 'clock' pulse train of known frequency f_r and of sufficient stability for the conversion resolution contemplated.

It is convenient to regard a signal frequency f as representing a numerical quantity m that lies in the range $0 \rightarrow M$, where M is an integer. In this convention, M is chosen to indicate the desired resolution according to:

'per unit' resolution = 1/M

so that m need be known only to the nearest integer,

$$f = B + F.\frac{m}{M} \tag{3-65}$$

where F is the range of f and B is a bias frequency.

Pulse-frequency, as a measurement variable, has the inconvenient property of not being continuously observable. Pulse-frequency manifests itself (as a retrospective average) only at intervals which themselves depend on its value. If the time-continuous variable (f or m) is to be reconstructed from observations of the instant of occurrence of each pulse, then a form of sampling theorem must be satisfied. In effect the manner and rate at which f changes must be constrained. f as a function of time must not contain significant spectral components with frequencies above half of f itself.

Certain types of frequency-to-number converters require a further constraint to be satisfied. This is that the frequencies obtained by measuring adjacent pulse-intervals should not differ by more than the specified frequency resolution; the corresponding values of n should not differ by more than 1.

That is

$$f_i \sim f_{i+1} = \frac{1}{f}\left|\frac{df}{dt}\right| < \frac{F}{M}$$

or

$$\left|\frac{d(\log f)}{dt}\right| < \frac{F}{M} \quad . \qquad (3\text{-}66)$$

The constraint of equation (3-66) is more severe than the previous one, and within the limits of accuracy imposed by rounding, effectively includes the previous one. Provided (3-66) is satisfied, as it will generally be, a number of frequency-to-number conversion techniques are available. The following will be discussed:

period measurement (+ arithmetic inversion)

pulse-counting over a fixed time

closed-loop frequency tracking

successive comparison with known frequencies

3.7.1 Period measurement.

Two successive signal pulses are used to respectively open and shut a gate, through which reference frequency pulses have access to a counter. The number of pulses reaching the counter will be a measure of the ratio of signal to reference pulse periods.

If 'a' pulses reach the counter, then using equation (3-65)

$$a \doteq \frac{1}{f}.f_r = \frac{M}{F.m + B.M}.f_r$$

As 'a' is rounded to an integer, to meet the requirement that m be known to the nearest integer entails that a unit change in m causes at least a unit change in 'a' :

$$\left|\frac{da}{dm}\right| \geqslant 1 \qquad \text{for all m in the range } 0 \rightarrow M .$$

i.e
$$f_r \geqslant F\,M\left(\frac{m}{M} + \frac{B}{F}\right)^2 .$$

To obtain sufficient resolution, the reference frequency must be high enough to satisfy the inequality above even when 'm' has its maximum value of M. That is:

$$f_r \geqslant F.M\,(1 + \frac{B}{F})^2 . \tag{3-67}$$

While f_r should be high to obtain sufficient period resolution for signal frequency (f) a maximum, a high reference frequency will result in large values of the count 'a' when the signal frequency is low. The counter used for 'a' may have to be very large. As $a_{max} \doteqdot f_r/B$, substitution of (3-67) gives

$$a_{max} \geqslant M\,(1 + \frac{B}{F})(1 + \frac{F}{B}) , \tag{3-68}$$

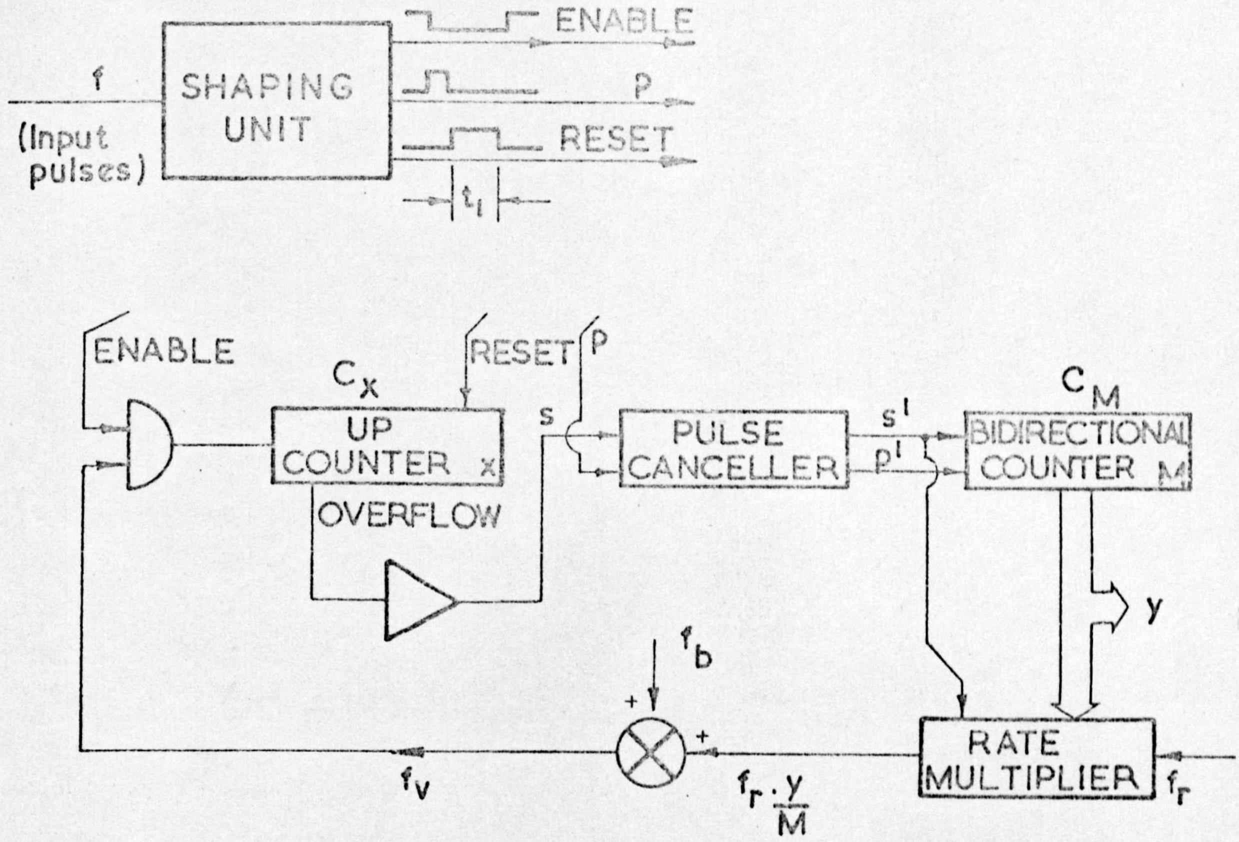
which for any B gives

$$a_{max} > 4M .$$

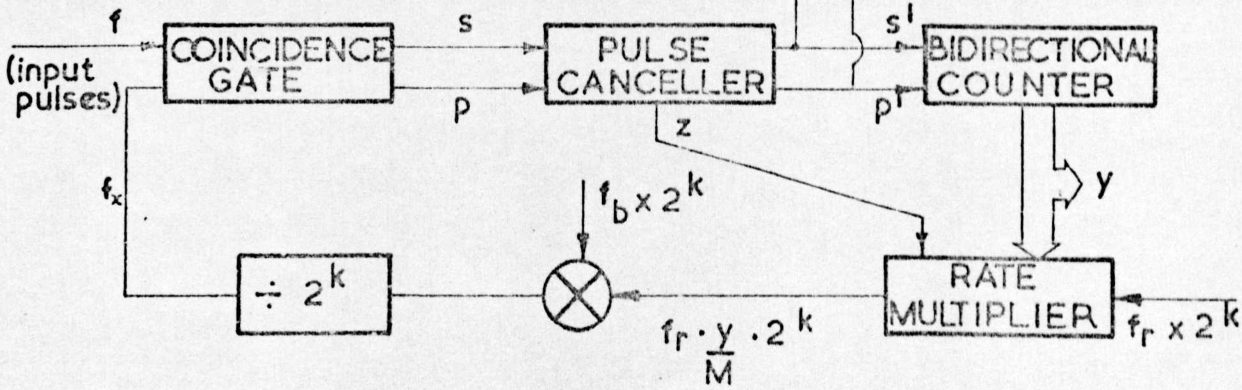
When 'm' is encoded into signal frequency 'f' without any bias, the equations above are unhelpful, giving an infinite value for a_{max}. A practical converter under such conditions may be constructed to saturate at

$$a_{max} = M.a_{min} = M.\frac{f_r}{F} \geqslant M^2 .$$

An example: The signal frequency lies in the relatively low band 1 kHz $< f <$ 2 kHz , and demanded resolution is 0.1% (M = 1000).

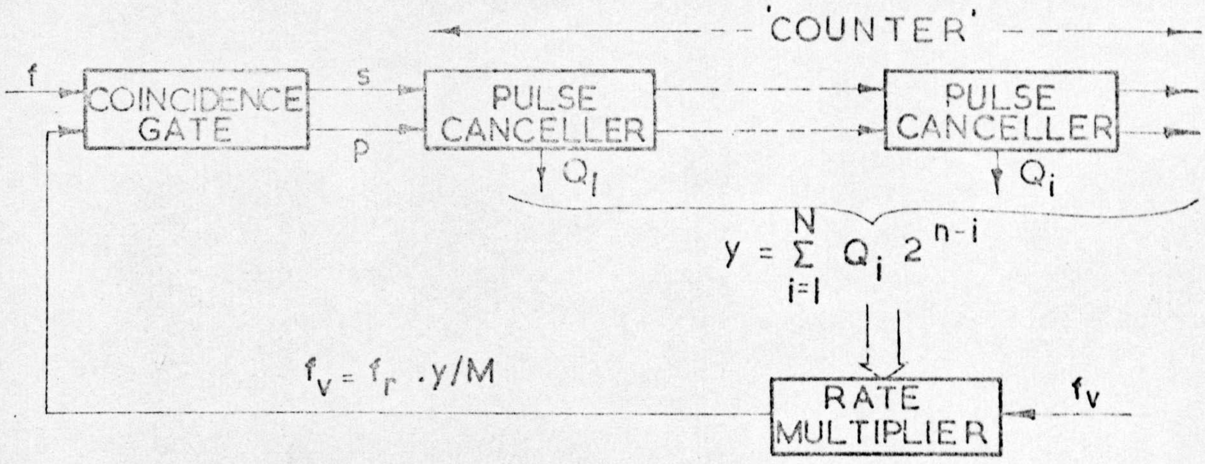


(a) Period tracking circuit, $y \propto$ (f - const.)

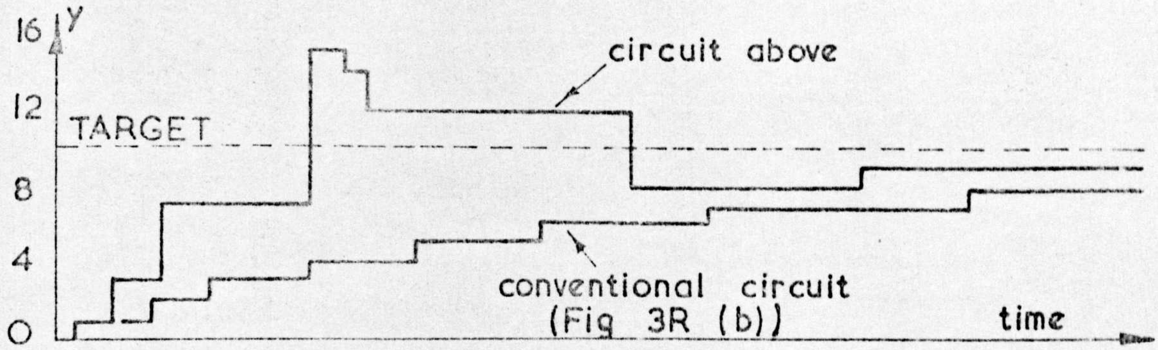


(b) Frequency tracking circuit, $y \propto$ (f - const.)

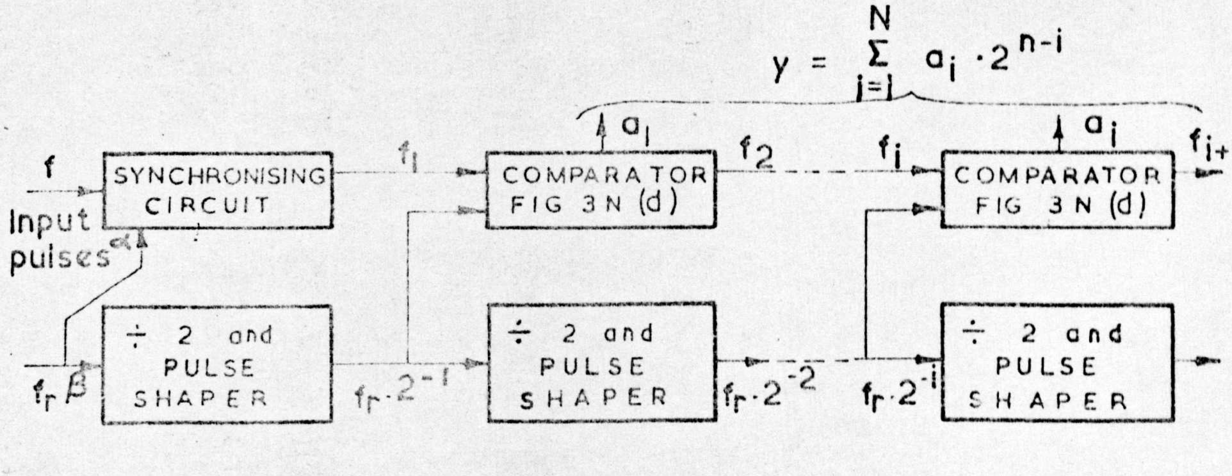
FIG. 3R Frequency-to-number conversion circuits.



(a) Frequency tracking using 'counter' of pulse - cancellers



(b) Step responses, $f = 0 \rightarrow f_r \cdot 5/8$



(c) Frequency tracking using successive comparison

FIG. 3 S Further frequency - to - number conversion circuits

Then (3-67) and (3-68) indicate that:

reference frequency > 4 MHz

maximum count > 4000.

The disadvantage of a high reference frequency is that the counting elements must be high speed. The disadvantage of a high maximum count (a_{max}) is that a large counter is required and that arithmetic inversion (to obtain frequency from period) must be applied to large numbers.

There is a closed-loop configuration based on period measurement that also achieves arithmetic inversion. A schematic diagram is shown in Fig. 3R(a). The incoming signal pulses gate an internal pulse-train f_v into a counter C_x to give a count c. Hence

$$c = f_v / f.$$

The counter C_x has a maximum capacity X, so it overflows for every count during which c exceeds X. The overflow of this counter is used to modify the rate of the internal pulse-train. This feedback results in a tracking action that endeavours to attain the equilibrium state:

$$X = f_v / f, \quad \text{but} \quad f_v = f_b + f_r . \frac{y}{M} ,$$

so

$$y = M.X\, f/f_r - M\, f_b / f_r \tag{3-69}$$

Equilibrium state can be maintained provided condition (3-66) is satisfied. As from (3-69) and (3-65):

$$y = K_1 f - K_2 \qquad \text{and} \qquad f = K_3 m + K_4$$

where

$$K_1 = MX/f_r , \quad K_2 = Mf_b / f_r , \quad K_3 = F/M , \quad K_4 = B.$$

By suitable choice of X, f_r and f_b it is possible to obtain the correct conversion (final output = original input):

$$y \doteqdot m \qquad (3\text{-}70)$$

Equation (3-69) and its dependents contain an 'approximately equals' sign. This reflects a practical constraint on the conversion method. The analysis above assumes that on arrival of a signal pulse the following actions are performed in negligible time: (i) the gate controlling entry to counter C_x is closed, (ii) the state of the overflow bit of counter C_x is shifted into the pulse canceller, (iii) counter C_x is reset to zero, (iv) the entry gate is re-opened. In practice these actions take time, as shown in the timing waveforms at the top of the diagram. The signal period observed will be effectively t_1 short of its true value: $1/f$; this modification to the theory can be neglected provided $f.t_1 < 1/M$. Otherwise, and allowing for rounding, equation (3-69) should be written

$$y = K_1 \frac{f}{1 + f.t_1} - K_2 \pm 1.$$

The period tracking circuit just described appears to be a novel one, possibly useful in decoding fairly low pulse-rates. A number of refinements are possible, such as those included in Fig.R(a) to reduce fluctuations in the numerical output: y.

3.7.2 Pulse-counting.

Counting pulses for a fixed time is the most widely used method of frequency-to-number conversion. It is employed in commercial timer-counters, in certain digital voltmeters, medicine (heartbeats in one minute), digital tachometers etc. It is a sampling technique, and for reliable conversion of a varying frequency, f(t), the sampling theorem should be satisfied. For a given pattern of frequency variation, the (Shannon's) sampling theorem places an upper bound T_2 upon the count time T.

$T_2 = 1/2a$ where a is the frequency of the highest significant spectral component of f.

Conversely the demands of resolution place a lower bound T_1 on T, for the larger T (given : f), the larger the numerical value obtained.

$T_1 = M/F$ where F is the range of f and 1/M is the fractional resolution.

For adequate waveform reconstruction and adequate resolution, $T_1 < T < T_2$; so a suitable conversion period is only available if $T_1 < T_2$.

The definition of T_2 above depends upon the force given to the phrase 'significant spectral component'. In the knowledge that only a finite resolution is required, T_2 can be rephrased as "that counting period just short enough to allow the frequency's waveform to be reconstructed to an accuracy of F/M." Thus relaxing the fractional resolution 1/M lowers T_1 and raises T_2. It is henceforth assumed that M has been appropriately chosen so that $T_1 < T_2$ and that T satisfies $T_1 < T < T_2$.

The output of a timer-counter convertor that obtains a number $y(t)$ from a frequency $f(t)$ is investigated in [332], viz:

$$y(t) = \left[\int_{t-T}^{t} f(t)dt\right] \text{ passed through a sample-and-hold.}$$

Assuming the sampling theorem to be satisfied by appropriate choice of T, them neglecting high frequency components in $y(t)$, a transfer function can be introduced

$$\frac{Y(s)}{F(s)} = \frac{(1-e^{-sT})}{s}.\frac{(1-e^{-sT})}{sT} = T.e^{-sT}\left(\frac{e^{\frac{sT}{2}} - e^{\frac{-sT}{2}}}{s\,T}\right)^2. \quad (3\text{-}72)$$

The conversion is thus associated with a gain of T, a delay of T, and a distortion factor $D(s)$. Writing $a(x) = \sin(x)/x$,

$$D(j\omega) = a^2\left(\frac{\omega T}{2}\right) \qquad \text{where } \omega \text{ is the pulsatance of a component of } f(t).$$

(3-73)

If the error introduced during the conversion (neglecting quantisation) is $e(t) = y(t) - T.f(t)$, then

$$\frac{E(j\omega)}{T.F(j\omega)} = -(j\omega T) - \frac{7}{12}(\omega T)^2 + j(\omega T)^3 + \frac{31}{360}(\omega T)^4 + \ldots..$$

This error-to-signal power ratio at the frequencies of main interest (for which $\omega T < 1$) approximates to:

$$\left|\frac{E(j\omega)}{T.F(j\omega)}\right|^2 = \text{Power ratio} = (\omega T)^2\left(1 - 0.16(\omega T)^2 + 0(\omega T)^4\right)$$

So those components of $f(t)$ with pulsatance ω approaching $1/T$ will be substantially distorted during conversion to number $y(t)$. This distortion is primarily due to the average delay introduced during conversion.

3.7.3 Frequency tracking

Figure 3R(b) shows the block diagram of a frequency tracking circuit. The feedbank is arranged so that the frequency : f_v of an internally generated pulse train tends towards the frequency f of the incoming pulse train. As f_v is linearly related to the number y, this number tends towards some multiple of f. Circuits of this general type are described by Lundh,[301] Wood[327] and Leonhard[305].

The circuit of Fig. 3R(b) possesses three refinements of the most basic arrangement, viz.

(i) There is a means of effectively removing a bias in the incoming frequency, before conversion to number : y.

(ii) The technique of driving the scaling circuit with a high clock frequency, and subsequently dividing by an integer (÷A), reduces jitter in the locally generated pulse train (f_v).

(iii) Normal operation of the feedbank would result in a 1 bit flutter in output number y when the input frequency was steady. The link 'z' on the diagram is a 1 bit by-pass to the counter containing y and removes flutter in y. This effect was found to hold experimentally, even when the input pulse train (f) exhibited considerable jitter.

The full equations of operation of the circuit are as follows:

$$\text{error rate} = f_e = f - f_v$$

$$\text{count} = y = \int^t f_e \, dt \qquad (3\text{-}74)$$

$$\text{feedback} = f_v = y.\frac{f_r}{M} + f_b = \frac{y}{T} + f_b$$

$$\text{where} \quad T = M/f_r$$

Expressing equations (3-74) using Laplace Transforms:

$$Y(s) = T.\frac{F(s)}{1 + sT} - \frac{T f_b}{s(1 + sT)}$$

or introducing the notation $f_1(t) = f(t) - f_b$:

$$Y(s) = T.\frac{F_1(s)}{1 + sT} \tag{3-75}$$

$$E(s) = Y(s) - T.F_1(s) = \frac{-sT^2.}{1 + sT} F(s) \tag{3-76}$$

$$\left.\frac{\text{Error power}}{\text{Signal power}}\right|_{\text{at}\,\omega} = \left|\frac{E(j\omega)}{T.F(j\omega)}\right|^2 = \frac{(\omega T)^2}{1 + (\omega T)^2} \tag{3-77}$$

A variation on the conventional frequency-to-number tracking converter may be obtained by replacing the conventional bidirectional counter in Fig. 3R(b) by the special 'counter' of cascaded pulse-cancellers discussed in section 3.5. Such an arrangement is shown in Fig. 3S (a). It will be recalled that a counter of pulse-cancellers will on average take much less time than a conventional counter to execute large numerical changes, but is liable to excessive overshooting while executing small changes. These effects result in oscillatory outputs when such a counter is part of a closed loop. Fig. 3S (b) compares the response of tracking loops containing conventional and special counters to a step change in input frequency. The special counter gives an output number that approaches its final value more rapidly (often much more rapidly) than obtains with a conventional counter. However in most applications, rapid settling of a frequency-to-number converter following step changes in frequency is of less significance than the ability to follow, smoothly, slower frequency variations. In this latter rôle the converter employing the conventional counter performs better.

3.7.4 Successive comparison.

A frequency-to-number convertor based on successive frequency comparison has been proposed by Martin[330]. This convertor is a near relative to that described at the end of the previous section. It gives a numerical output that continuously follows the frequency of the incoming pulse train, so it is not a sampling converter. However it is not possible to represent its action by a linear transfer function; the response following any change in input frequency depends partly upon input phase (relative to that of the reference pulse train).

The working of this converter can be deduced from Fig. 3S (c). The incoming frequency is compared with $f_r/2$; if $f > f_r/2$, a pulse train of rate $(f - f_r/2)$ is passed on to the next comparator; if $f < f_r/2$, the whole pulse train (of rate f) is passed on. The next comparator repeats the process using $f_r/4$ as a standard. Unlike a voltage comparator, however, a frequency comparator may take a considerable time to effect the comparison. A maximum value for this time is 1/(frequency difference), and an average value is 1/2(frequency difference). Depending on the initial phase relationship between two pulse trains, the comparison of their frequencies can take from 0 to the maximum value just described.

Martin, on the apparently false assumption that a frequency comparator responds in not more than the reciprocal of <u>one</u> of the frequencies fed to it, finds that a comparator type frequency-to-number converter will resolve an incoming frequency to an accuracy of f_r/M in time not greater than $2M/f_r$. His conclusion, (despite his premise) seems substantially correct, as may be confirmed by considering changes of the sort most unfavourable to the system, viz. from $f = (0.0111...11)f_r$ to $f = (0.1000.....01)f_r$ which the output number takes not more than $2\frac{1}{2}M/f_r$ to follow.

The maximum response time of this converter, to a frequency change of any amplitude, is about $2M/f_r$; the average response time is about M/f_r. Thus the converter has a speed of response similar to that of the pulse-counting converter of Section 3.7.2 without the disadvantages of sampling. The converter has the severe disadvantage of occasional large overshoots, albeit of short duration, every time a high order bit (of the output number) changes.

If the input frequency f ranges over the full range f_b to $f_b + f_r$, the energy (integral of amplitude squared) of the output error is approximately: $M^2 n/2f_r$. This assumes the range of the output number to be $M = 2^n$. When the input frequency traverses its range f_r in time L, the mean output error power will be : $M^2 n/2f_r L$. If the input frequency only traverses a fraction λ of the full range f_r in time L, the mean output error power can be shown to equal approximately $M^2 \lambda \log_2(\lambda M)/2f_r L$.

For the particular case: $f = \text{const.} + A.f_r \cos\omega t$, where $A < \frac{1}{2}$:

Defining $K = 2MA$, $T = M/f_r$ and if ωT not $\gg 1$

$$\text{Error power} \doteqdot \frac{M^2 .2A \log_2(2MA)}{2f_r . \pi/\omega} = \frac{(\omega T)K \log_2 K}{2\pi} \tag{3-78}$$

$$\frac{\text{Error power}}{\text{Signal power}} \doteqdot \frac{(\omega T)K \log_2 K}{2\pi . \frac{1}{2}M^2 A^2} = 1.27(\omega T) \frac{\log_2 K}{K} \tag{3-79}$$

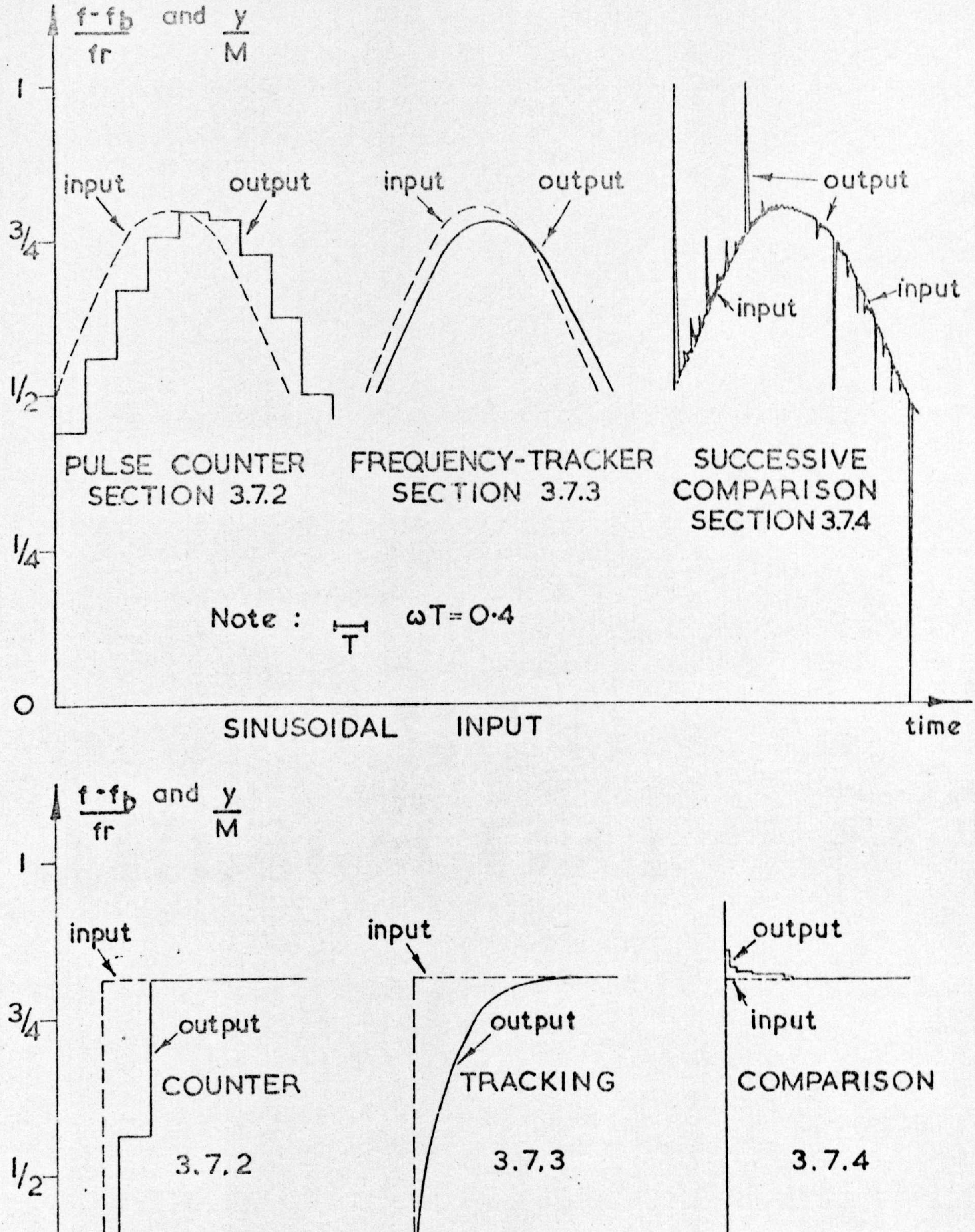


FIG 3T RESPONSES OF FREQUENCY-TO-NUMBER CONVERTERS TO SINUSOIDAL AND STEP FREQUENCY CHANGES.

(Normalised input frequency $(f-f_b)/fr$
Normalised output number y/M)

3.7.5 Comparison of converter types.

The four types of converter considered in sections 3.7.1 to 3.7.4 have somewhat different characteristics. The general nature of these is visible in Fig. 3T where the responses to sinusoids and steps are shown for three of the converter types.

(i) The period-tracking designs of section 3.7.1 are only applicable when the input frequency f is low. For a numerical resolution into M parts, internal frequencies of at least $4Mf_{max}$ and as high as M^2f_{max} are required in the converter. Counting circuits that will work with pulse-rates in excess of 20MHz are expensive and difficult to use. The period-tracking design also introduces a slight non-linearity into a frequency-to-number converter, if the frequency is too high.

Where frequencies are low (e.g. $<$100Hz to$<$5kHz depending on resolution), this type of converter is markedly superior to other types, responding with delays only a fraction of those experienced with the others.

(ii) Frequency-to-number counters of section 3.7.2 are simple to design, but have fairly high errors (Equation 3-73) and may cause alliasing due to their sampling action. The averaging inherent to the conversion process is a useful property when the incoming pulse train shows much jitter. However the output number will show some flutter, varying by 1 even when the input pulses are equally spaced.

This type of converter is useful where the numerical output undergoes further processing in a digital device, or where periodic digital records are to be kept, (e.g. data logging).

(iii) The tracking converters of section 3.7.3 are most useful when the numerical output is used to control an analogue device, or where rate information is required. The continuous output responds quickly (although not very accurately) to changes in input. This sort of converter gives very slightly lower errors, and fewer stabilisation problems (than the counting converter above) when it forms part of a feedback control loop.

(iv) The frequency-comparator design of section 3.7.4 gives an output subject to large but brief fluctuations. The mean squared error, while tracking a changing input, is (Equation 3-79) a non-linear function of input amplitude. In general, however, the tracking errors are less than are experienced with the last two types. This somewhat complex converter could form part of a very efficient frequency-to-voltage converter and would be valuable in other applications where its momentary aberrations could be disregarded or filtered out.

Table of comparison : frequency-to-number converters.
(See also Fig. T)

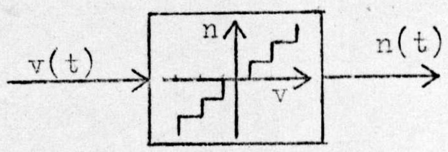
Type of converter	Pulse Counter	Tracking	Successive Comparison
Section	3.7.2	3.7.3	3.7.4
Error Power/ Signal Power	$(\omega T)^2(1 - 0.16(\omega T)^2 \ldots)$	$\dfrac{(\omega T)^2}{1 + (\omega T)^2}$	$1.27(\omega T)\dfrac{\log_2(2MA)}{2MA}$
Power ratio for A=0.25, T=0.5	0.240 (SNR=6.1db)	0.200 (SNR=7.0db)	0.011* (SNR=19.4db)
Power ratio for A=0.062, T=0.1	0.010 (SNR=20db)	0.010 (SNR=20db)	0.007* (SNR=21.6db)

*M assumed as 1024

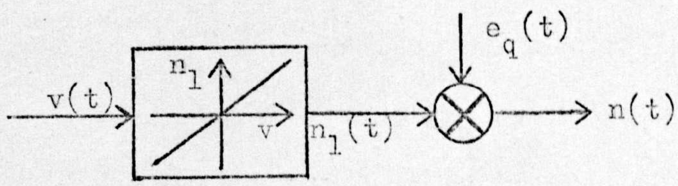
3.7.6 Effects of rounding

All analogue-to-digital conversions and some wholly digital operations are subject to the effects of rounding, due to the finite length of the registers employed. There is an extensive literature about rounding or quantisation, the greater part (e.g.340) being concerned with the influence of word length upon the results of complex arithmetical operations in computers. However a number of papers have been written concerning quantisation arising during analogue-to-digital conversion[333-339]. The conversions discussed (excepting Nightingale and Richards[339]) have involved continuous and continuously observable analogue variables, and it is convenient here to identify such conversions as 'voltage-to-number' conversions.

In well designed apparatus, quantisation errors will generally be small. As their calculation is usually complex, it is desirable to account for them by statistical techniques. To a moderate level of accuracy, quantisation errors can often be represented by a small random signal that is added to the output signal from the device where rounding occurs. For example an actual voltage-to-number converter:



could be replaced by an 'ideal' convertor and a rounding-noise source:



This allows an initial analysis of a system that ignores rounding, followed by superposition of the variations due to the injection of smallrandom signals at appropriate points. For this sort of analysis both first and second order statistics of the error 'noise' are required.

Widrow[333] examined sampling voltage-to-number conversion, and established that for a voltage varying over a range of several quan-tisation steps (each of amplitude 'q' volts), the quantisation error had the first order statistics:

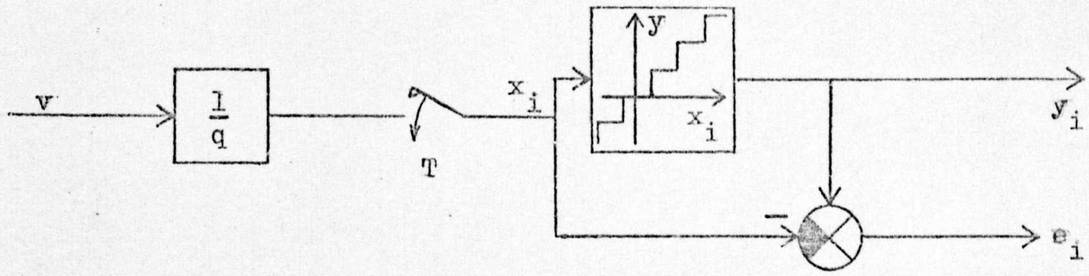
$$\begin{aligned} p_e(x) &= 1/q \quad \text{when} \quad -q/2 \leqslant x < q/2 \\ p_e(x) &= 0 \quad \text{when} \quad |x| \nless q/2 \\ \langle e_q(t) \rangle &= 0 \\ \langle e_q^2(t) \rangle &= \sigma_e^2 = q^2/12 \end{aligned} \tag{3-80}$$

He further established that where the quantisation steps were small enough, and the voltage sufficiently slowly varying, so that the amplitude probability density of the voltage could be fully reconstructed from the amplitude probability density of the output numbers (quantised in time and space) - then the quantisation errors could be regarded as samples from white noise. Successive quantisation errors would be uncorrelated. Although it is difficult to determine (in practice) whether Widrow's criterion holds, methods based on it are so much more manageable than any alternatives as to justify their examination in all cases. A rule of thumb, that is sufficiently accurate for most applications, is: use Widrow's approach if the average number of quantisation steps traversed by the voltage between successive conversions exceeds 2.

A quite different approach[336], valid where the voltages enter-ing voltage-to-number converters in a system are almost static, is to locate 'dead bands' due to rounding, over which the system gain is effectively zero. With 2 state digital devices, this approach is particularily relevant.

Frequency-to-number conversion differs somewhat from voltage-to-number conversion. The number obtained is a function of both phase and frequency of the incoming signal, although frequency predominates. The analysis below relates to the conventional sampling frequency-to-number converter of section 3.7.2 ; its results give some indication of effects to be expected in other types of converter.

Widrow's analysis was applied to the system:



If x_i is written as $X_i + r_i$, where X_i is integer and $0 \leqslant r_i < 1$,

then when $r_i \geqslant \frac{1}{2}$: $y_i = X_i + 1$ and $e_i = 1 - r_i$,

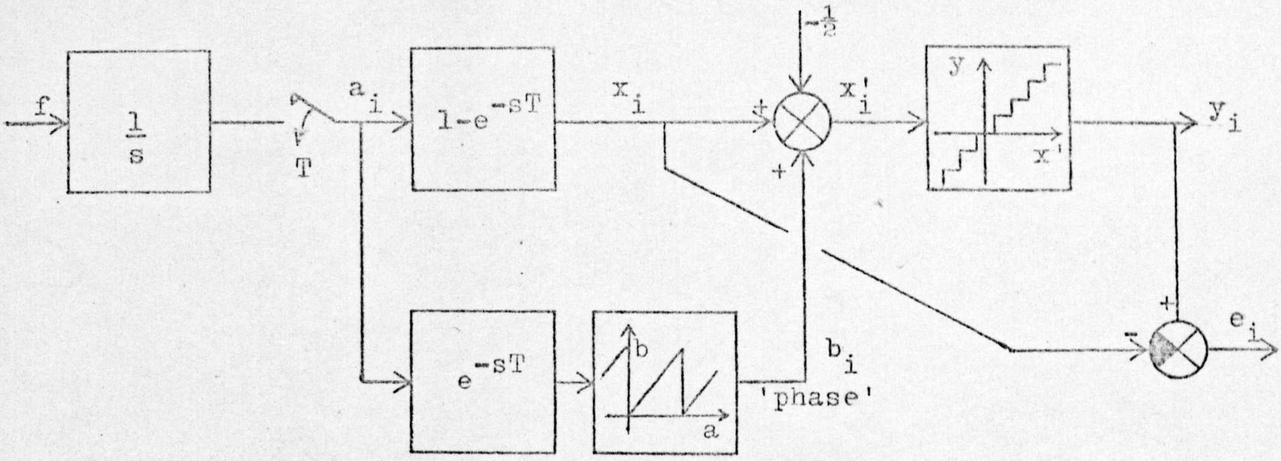
but when $r_i < \frac{1}{2}$: $y_i = X_i$ and $e_i = -r_i$.

Thus the rounding error lies in the range $-\frac{1}{2}$ to $+\frac{1}{2}$ unit. If the amplitude distribution of the change $(x_{i+1} - x_i)$ between samples is smooth and spans at least ± 2 units, the distribution of the error will be approximately uniform in the range $-\frac{1}{2} < e < \frac{1}{2}$. Moreover this distribution does not significantly depend upon the error of the previous conversion, i.e.

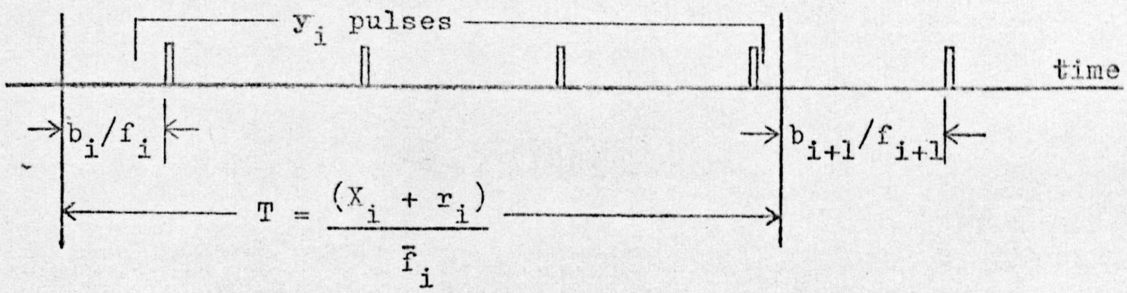
$$p_{e_i}(k;\, e_{i-1}) = p_{e_i}(k) = 1, \text{ if } |k| < \tfrac{1}{2}$$
$$= 0, \text{ if } |k| \geqslant \tfrac{1}{2}$$

This shows that successive conversion errors are nearly independent.

Frequency-to-number conversion may be represented by the block diagram:



corresponding to the waveform:



The block diagram above differs from that of Widrow's analysis in two respects. Firstly x_i is related to f by not only a conversion constant T, but also a running-averager with transfer function $(1 - e^{-sT})/sT$. Secondly there is the influence of the 'phase' b_i. The first difference has negligible effect upon quantisation error statistics, as the amplitude distribution of x has much the same general shape as that of f, sufficient for the arguments of error independence still to hold. The second difference is significant, and will now be shown to modify the amplitude distribution and independence of the errors.

Notation for the ensuing analysis:

j,k,l,m	are	dummy variables
b	is	phase at beginning of i^{th} conversion
x	is	true output " " "
X	is	integer part of x
r	is	non integer part of x, so $x = X+r$; $0 \leqslant r < 1$
y	is	actual output of i^{th} conversion
e	is	output error " " "

b', x' etc, are corresponding values for $(i+1)^{th}$ conversion.

b and r may be assumed evenly distributed across their ranges, and independent of each other. Under conditions where Widrow's arguments hold, r' is effectively independent of r.

So if $0 \leqslant k < 1$, $\quad p(k)=p_b(k)=p_r(k)=p_{r'}(k)=p_r(k;b)=p_{r'}(k;r) = 1$,

if $k<0$ or $k \geqslant 1$, $p(k)=0$.

$$(3\text{-}81)$$

From the diagram above:

$$\text{If } r > b \quad \text{then} \quad y = X+1, \quad e = 1-r, \quad b' = 1+b-r = b+e,$$

$$\text{also} \quad 0 < e < 1-b \qquad (3\text{-}82)$$

$$\text{If } r < b \quad \text{then} \quad y = X, \quad e = -r, \quad b' = b-r = b+e,$$

$$\text{also} \quad -b < e < 0 \qquad (3\text{-}83)$$

Equations (3-81), (3-82) and (3-83) will now be used to determine various error statistics, and interdependences.

From (3-81) and (3-82)

$$\begin{aligned} p_e(k;r>b;b=j) = p_r(1-k;r>b;b=j) &= 1/(1-j) \quad \text{if } 0 \leqslant k < 1-j \ , \\ &= 0 \qquad\quad \text{if } k<0 \ \text{ or } \ k \geqslant 1-j \ . \end{aligned}$$

From (3-81) and (3-83)

$$\begin{aligned} p_e(k;r<b;b=j) = p_r(-k;r<b;b=j) &= 1/j \qquad \text{if } -j<k \leqslant 0 \ , \\ &= 0 \qquad\quad \text{if } k \leqslant -j \ \text{ or } k>0 \ . \end{aligned}$$

So

$$\begin{aligned} p_e(k;b=j) = \ & p_e(k;r>b;b=j).P(r>b;b=j) \\ & + p_e(k;r<b;b=j).P(r<b;b=j), \end{aligned}$$

and as $P(r>j) = 1-j$ and $P(r<j) = j$,

then

$$\begin{aligned} p_e(k;b=j) &= 1 \quad \text{if} \quad -j<k<1-j \qquad (\text{i.e. } -k>j>1-k) \\ &= 0 \quad \text{if} \quad k \leqslant -j \ \text{ or } \ k \geqslant (1-j) \qquad (3\text{-}84) \end{aligned}$$

The general distribution of e is then

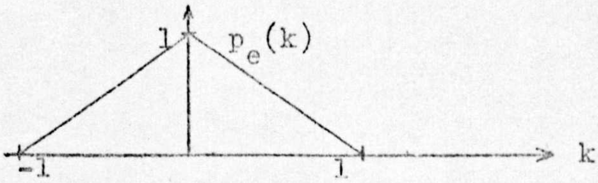
$$p_e(k) = \int_{j=-\infty}^{\infty} p_e(k;b=j).p_b(j).dj$$

$$= \int_{j=0}^{1-k} 1.1.dj \quad = \quad 1-k \qquad\qquad \text{if } 1 > k \geqslant 0$$

$$= \int_{j=-k}^{1} 1.1.dj \quad = \quad 1+k \qquad\qquad \text{if } -1 < k \leqslant 0 \tag{3-85}$$

So

$$p_e(k) = 1-|k|$$



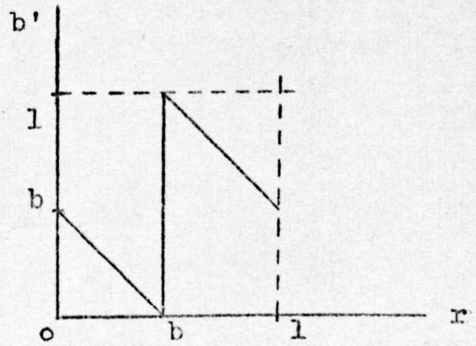
and the first order error statistics are:

$$\langle e \rangle = \int_{k=-1}^{1} k(1-|k|)dk = 0$$

$$\langle e^2 \rangle = \sigma_e^2 = \int_{k=-1}^{1} k^2(1-|k|)dk = 1/6 \tag{3-86}$$

The (normalised) variance for voltage-to-number conversion was shown (Eq. (3-80)) to be 1/12, so <u>frequency</u>-to-number error variance is twice that for voltage-to-number error.

The phase at the beginning of a conversion is apparently influenced (equations (3-82),(3-83)) by the phase and error of the previous conversion. This is deceptive, as it transpires that b' is independent of b, but not of e.



From equations (3-82) and (3-83) the relationship between b' and r depends on b.

An analysis similar to that leading up to Equation (3-84) gives:

$$p_{b'}(k;r>b;b=j) = 1/(1-j) \qquad \text{if } j \leqslant k < 1$$
$$= 0 \qquad \text{if } k < j \text{ or } k \geqslant 1 ,$$
$$p_{b'}(k;r<b;b=j) = 1/j \qquad \text{if } 0 \leqslant k < j$$
$$= 0 \qquad \text{if } k < 0 \text{ or } k \geqslant j .$$

As

$$P(r>b;b=j) = 1-j \quad \text{and} \quad P(r<b;b=j) = j ,$$

so

$$p_{b'}(k;b=j) = 1 \quad \text{where} \quad 0 \leqslant k < 1 \qquad (3\text{-}87)$$

Comparing (3-87) with (3-81) shows that $p_{b'}(k;b=j) = p_b(k)$, i.e. b' is independent of b.

The statistical relationship between the conversion error and b (the initial phase) will be identical to that between the error and the final phase. So as final phase of one conversion is the complement of the initial phase of the next conversion, (b'), equation (3-84) can be extended to relate both e and e' to b'.

As

$$p_e(k;b=j) = 1 \qquad \text{over} \quad -j < k < 1-j$$
$$= 0 \qquad \text{otherwise,}$$

so

$$p_e(m;b'=1-j) = 1 \qquad \text{over} \quad -j < m < 1-j$$
$$= 0 \qquad \text{otherwise,}$$

giving

$$p_e(m;b'=j) = 1 \qquad \text{over } j-1 < m < j. \qquad (3\text{-}88)$$

Equation (3-85) shows that $p_e(m) = 1-|m|$ over $-1 < m < 1$, so $p_e(m,b'=j)$ differs from $p_e(m)$, and there is interdependence between e and b', i.e. between one conversion error and the following initial phase. As e' and b' are not independent, there is the possibility that e and e' are also not independent, as turns out to be the case.

The correlation between e and e' can be developed by choosing a value (say 'j') for the phase b'.

If $b'=j$, then $\langle e,e'\rangle = \int_{m=-\infty}^{\infty}\int_{k=-\infty}^{\infty} m.k.p_e(m;b'=j)\ p_{e'}(k;b'=j)\ dm\ dk$,

$$= \int_{m=j-1}^{j} m.1.dm \quad x \quad \int_{k=-j}^{1-j} k.1.dk,$$

$$= -j^2 - j + \tfrac{1}{4}.$$

Now allowing b' to vary over its range

$$\langle e.e'\rangle = \int_{j=-\infty}^{\infty} dj.p_{b'}(j).\ \langle e.e'\rangle_{\text{given } b'=j}$$

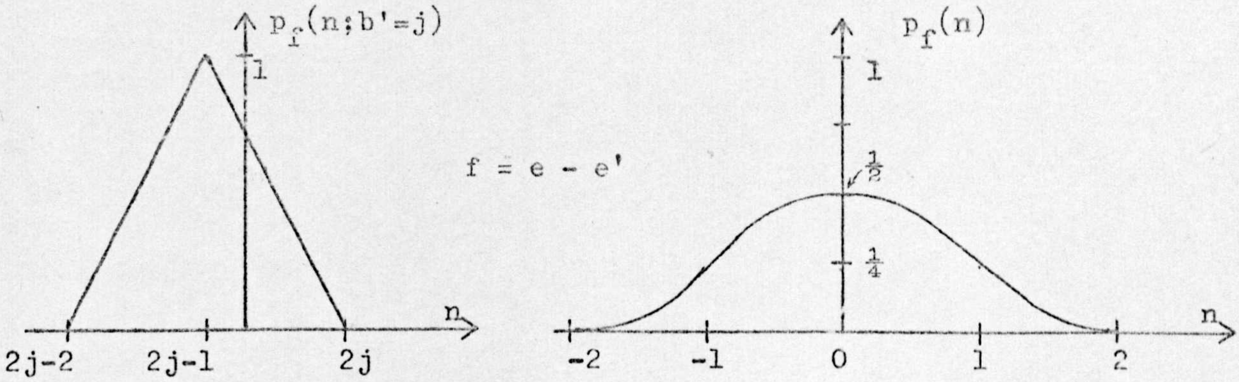
$$= \int_{j=0}^{1} (-j^2 - j + \tfrac{1}{4})dj$$

$$= \frac{-1}{12} = \frac{-1}{2}\sigma_e^2 \qquad (3\text{-}89)$$

There is some interest in the error ($f = e - e'$) in the output of a device that performs two successive frequency-to-number conversions and then subtracts one number from the other. Such a device will be discussed in the next section which concerns thetime-derivative of pulse-frequency.

By choosing b' (=j), the conditional distributions of e and e' can be so convoluted as to yield $p_f(n;b'=j)$. Then the final distribution $p_f(n) = \int_{j=0}^{1} p_f(n;b'=j)\ p_b(j)dj$.

The results of these operations are shown graphically below:



The correlation between successive errors (Equation (3-89)) can be used to calculate the variance of f = e - e'.

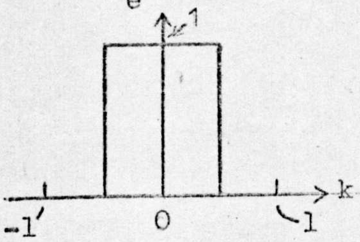
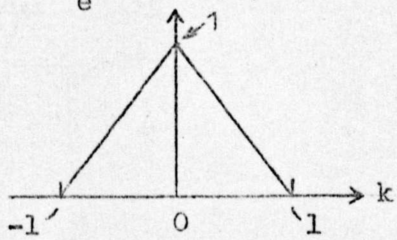
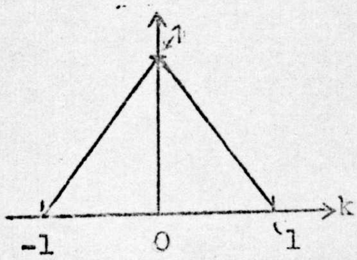
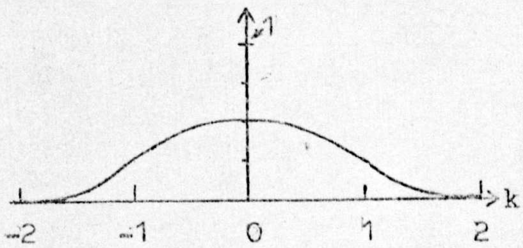
$$\langle (e - e')^2 \rangle = \langle (e')^2 \rangle - 2\langle e.e' \rangle + \langle e^2 \rangle$$

$$\sigma_f^2 = \frac{1}{6} + \frac{1}{6} + \frac{2}{12} = \frac{1}{2} \qquad (3\text{-}90)$$

Various experiments were performed, using a computer, to investigate the errors in converting a frequency (from a randomly varied signal generator) into numbers. Due to finite computer word length, (6 bits), these measured errors are themselves subject to quantisation errors, and the deduced statistics require some sort of Shepherd's factor correction.

Sample size (Experiment)	$\bar{e}$	σ_e^2	σ_f^2	$\langle e_i e_{i+1} \rangle$	$\langle e_i e_{i+2} \rangle$	$\langle e_i e_{i+3} \rangle$
1024	0.000	0.166	0.492	-0.090	0.011	-0.003
1024	0.000	0.167	0.492	-0.083	-0.004	0.007
1024	0.000	0.173	0.515	-0.089	0.003	-0.002
6 x 512	0.000	0.173	0.541	⟵ max.values observed		
		0.162	0.431	⟵ min.values observed		
Theory ⟶	0	0.167	0.5	-0.083	0	0

The experimental results support the theoretical predictions. In the table below the error characteristics of a frequency-to-number and a voltage-to-number converter are compared. The latter has the smaller errors.

Voltage-to-number	(Widrow's conditions satisfied)	Frequency-to-number
$\langle e \rangle = 0$	mean error	$\langle e \rangle = 0$
$\sigma_e^2 = 1/12$	error variance	$\sigma_e^2 = 1/6$
	error distribution $p_e(k)$	
$\langle e_i e_{i+1} \rangle = 0$	error correlation	$\langle e_i e_{i+1} \rangle = -1/12$
$\langle e_i e_{i+n} \rangle = 0$	$(n > 1)$	$\langle e_i e_{i+n} \rangle = 0$
$\sigma_f^2 = 1/6$	$(f = e_i - e_{i+1})$	$\sigma_f^2 = 1/2$
	difference error distribution $p_f(k)$	
$e_{max} = 1/2$	max. error	$e_{max} = 1$
$f_{max} = 1$	max. diff. error	$f_{max} = 2$

3.8. Differentiating a frequency

3.8.1 General considerations

In a number of circumstances, it is desirable to obtain a measure of the rate of change, f'(t) of a pulse-frequency, f(t). The measure may itself be a frequency or a number, although of course whichever form it takes, it can be converted to the alternative form. As well as the various pulse-techniques to be discussed here, there are also analogue differentiating techniques which can be applied if the incoming pulse frequency is first decoded into a voltage or other analogue signal form.

The most common situation wherein knowledge of rate of change of frequency is required is in the measurement of acceleration. Velocity, especially angular velocity, is readily measured with traducers giving a frequency-coded output (pulse tachometers). Acceleration is the rate of change of velocity. A particular problem arises when measurement of starting acceleration is to be made, because with velocity close to zero, the tachometer pulses are infrequent. Any estimate of acceleration cannot be updated more often than once per tachometer period, which at low speeds may be unacceptably long. One solution to this problem is to introduce a bias into the encoding of velocity into frequency; Stephenson [341] suggested how this could be done.

While a counter acts as an integrator to a pulse-frequency signal, there is no such simple device for differentiating a frequency. The discontinuous availability of information with pulse-frequency encoding and the likelihood of pulse-jitter limit the quality of any differential signal. For a jitter-free pulse-train, the best possible estimate of rate of pulse-rate change would be obtained by

(a) measuring successive pulse periods

(b) obtaining the reciprocal of each measurement, i.e. the corresponding frequencies

(c) differencing successive estimates of frequency, and multi--plying by the latter frequency of each pair.

This estimate is $f.\nabla\frac{1}{T} = \frac{1}{T}\Delta f \doteq \frac{df}{dt}$; its calculation involves substantial computation and hence costly measuring apparatus.

In the pages following, four methods of estimating a frequency derivative will be described that do not require parallel digital computations, and may be effected using relatively cheap apparatus. The four methods may be placed in two classes:

A. methods involving comparison of a past and a present measure of frequency.

B. methods involving the conversion of frequency into number by a tracking technique, the rate of tracking being a measure of the rate of change of frequency.

In order to discuss the limitations of the different methods, it is useful to postulate the form of the signal to be differentiated. The pulse frequency will therefore be assumed to have the general form $f(t) = f_b + f_o \sin \omega t$, where $\omega \ll f_o$ and where $f_b > f_o$.

3.8.2 Delay and subtract methods (Class A)

A pulse-train may be delayed and subtracted from itself. The result is a signed pulse-train (P & S or U & D convention), of effective frequency $f_r(t)$. Using ∇ operator to indicate change over time T, then:

$$f_r(t) = \nabla f(t) = f(t) - f(t - T)$$
$$= T\, f'(t) - \frac{T^2 f''(t)}{2!} + \frac{T^3 f'''(t)}{3!} - \text{etc.}$$

$$f_r(t) \doteq T\, f'(t) \left(1 - \frac{\nabla f'(t)}{2.f'(t)}\right). \qquad (3\text{-}91)$$

For small T, $f_r(t) \doteq T\, f'(t)$. If however a higher gain and hence longer T is used, the output $f_r(t)$ becomes proportional to a distorted form of the required derivative $f'(t)$. This distortion can be calculated for the proposed test signal $f(t) = f_b + f_o \sin \omega t$. Emplo ing Laplace transforms:

$$F_r(s) = (1 - e^{-sT})\, F(s)$$

$$= sT\, F(s).\, e^{\frac{-sT}{2}} .\, Sa(\frac{sT}{2}),$$

$$\text{where } Sa(x) = \frac{e^x - e^{-x}}{2x}.$$

So for the test signal

$$f_r(t) = \omega T.f_o \cos\, (\omega t - \frac{\omega T}{2}).\, \text{sinc}^{*}\, (\frac{\omega T}{2}) \qquad (3\text{-}92)$$

If the phase distortion is not to exceed say $\frac{1}{2}$ radian ($=28^o$) then

$$\omega T < 1.$$

If the amplitude distortion is not to exceed say 5% then

$$\text{sinc } T/2 > 0.95$$

$$\omega T < 1.1$$

The former condition is slightly more severe, so T should satisfy the condition

$$T < \frac{1}{\omega} \qquad (3\text{-}93)$$

Using the maximum acceptable value of T gives a derivative signal $f_r(t)$ of about the same amplitude as $f(t)$ neglecting the bias f_b.

The pulsatance ω of the test signal should relate to the dominant time constants of the system in which $f(t)$ occurs: so that as a rule-of-thumb, T should be chosen to roughly equal such time constants, but not exceed them. Thus T might vary from 10 milli-

* $\text{sinc}(x) = \frac{\sin(x)}{x}$

-seconds for a mechanical structure to 1 hour for a large chemical plant. There are no cheap and compact devices available that will delay a pulse train for such periods. The choice lies between magnetic recorders (e.g. disc or tape), shift registers and lumped-circuit delay lines. The first are bulky and rather costly; the second are still rather expensive considering that a storage of from 10,000 to 1,000,000 bits may be required; the third are not very stable. Stability of the delay $\underline{T}$ is necessary, for neglecting second order effects,

$$f_r(t) \doteq T f'(t) - T'f(t)$$

$$\text{where } T' = dT/dt$$

If $f_r(t)$ is only to reflect changes in $f(t)$, then

$$\left|T' f(t)\right| \ll \left|T f'(t)\right| .$$

As $T f'(t)$ is generally of smaller amplitude than $f(t)$

$$T' \doteq \frac{\nabla T}{T} < 0.02 \text{ ,say.} \qquad (3\text{-}94)$$

A lumped-circuit delay line, pneumatic or electronic, for which T of say ½ hour could be maintained constant within 2% over ½ hour, is not easily made.

The other member of this class of methods is the one most commonly used [320,332,342]. It yields a number nr that is nearly proportional to $f'(t)$. The number is obtained by a periodic calculation, and therefore possesses the usual disadvantages of a sampled variable. The basis of the method is to perform two frequency-to-number conversions at a fixed time spacing: the difference nr in the two conversions is then a measure of $f'(t)$.

The most direct method of implementing the method is to use a clock controlled bidirectional counter as follows:

(a) Set counter to zero

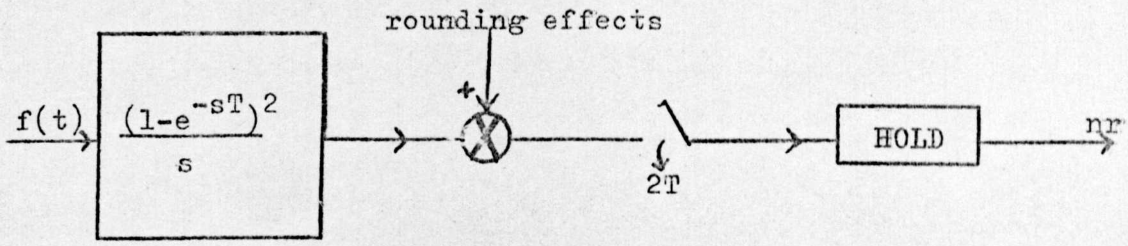
(b) Count up the signal pulses (of freq. $f(t)$) for time T

(c) Count down the signal pulses for time T

(d) Transfer the number left in the counter (nr) to an auxiliary register, and return to step (a).

$$nr = \int_{t-T}^{t} f(t)dt - \int_{t-2T}^{t-T} f(t)dt \qquad (3\text{-}95)$$

The method can be illustrated by the block diagram:



The method suffers from three forms of distortion. Firstly there is sampling distortion that introduces high frequency components into nr; these are partly attenuated by the 'hold' circuit. Secondly there is distortion that is frequency dependent and which increases with T. Thirdly there is distortion due to rounding, which (relatively) decreases with increasing T.

The low-frequency components of the output are related to the input by the transform relationship

$$NR(s) = sT^2 F(s). \text{ Factor } + \text{ rounding noise} \qquad (3\text{-}96)$$

where

$$\text{Factor} = e^{-2sT}. Sa^2(\frac{sT}{2}). Sa(sT)$$

For $f(t) = f_b + f_o \sin \omega t$, and for frequency dependent distortion not to exceed $\frac{1}{2}$ radian or 5% amplitude, as before, T should satisfy

$$\omega T < \tfrac{1}{2} \qquad \text{and} \quad \omega T < 0.45 \qquad \text{respectively.}$$

Taking the more severe condition

$$T < 0.45/\omega \tag{3-97}$$

The gain of this differentiator is T^2, insofar

$$nr \doteq T^2 f'(t)$$

This gain should be large enough for nr to be a sizeable integer, say 100 under conditions of maximum value for $f'(t)$. This is because (as discussed in section 3.6), nr is subject to rounding errors of up to ± 2 with a standard deviation of 0.7. So

$$T^2 f'(t)_{max} > 100$$

$$T^2 > \frac{100}{\omega f_o} \tag{3-98}$$

The constraints (3-97) and (3-98) are compatible provided

$$\frac{\omega}{f_o} < 0.002 \tag{3-99}$$

which is the condition already mentioned that the modulating (signal) frequencies must be much less than the carrier frequency.

3.8.3 Tracking error methods (Class B)

The simplest circuit to yield a derivative signal as a tracking error is that shown in Fig. 3R(b). Analysis yields the transform relationship

$$F_e(s) = \frac{sT}{1 + sT} \cdot F(s) \tag{3-100}$$

between the signed error frequency $f_e(t)$ and the input pulse-frequency $f(t)$. $T = M/f_r$ where M is the capacity of the bidirectional counter, and f_r is the reference frequency.

If $f_e(t) = T\,f'(t)$ when $f(t) = f_b + f_o \sin \omega t$ with a phase distortion of less than $\frac{1}{2}$ radian, then

$$\tan^{-1} \omega T < \tfrac{1}{2}\,, \qquad T < \frac{0.47}{\omega} \tag{3-101}$$

and with an amplitude distortion of less than 5%

$$(1 + \omega^2 T^2)^{-\frac{1}{2}} > 0.95 \qquad T < \frac{0.31}{\omega} \tag{3-102}$$

The more severe restriction is that of equation (3-102).

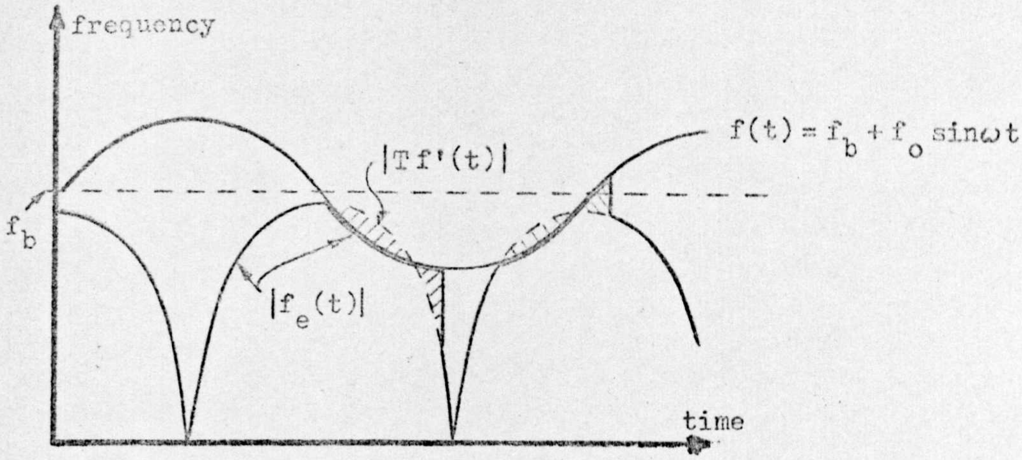
Alternatively the tracking circuit of Fig. 3R(a) may be used. Insofar this circuit tracks the incoming frequency within an error of 1 part in y, the tracking error is a very good approx--imation to $f'(t)$. (This error is the pulse-train P' interpreted by the sign S'). By increasing the capacity M of the bidirectional counter and its associated rate multiplier, any desired gain T can be obtained, where error frequency

$$f_e(t) = T\,f'(t) \tag{3-103}$$

However the tracking circuit is subject to the operational limit

$$\left| f_e(t) \right| < f(t) \tag{3-104}$$

So if $T \left| f'(t) \right| > f(t)$ there will be a distortion. This is illustrated in the sketch below. The shaded areas are equal.



For undistorted operation.

$$T < \frac{f_b - f_o}{\omega \ f_o} = \text{typically } \frac{2}{\omega} \qquad (3\text{-}105)$$

3.7.4 Comparison of methods

The primary characteristics of the four methods are set out below. 'Max. gain' indicates gain at which amplitude distortion approaches 5% or phase shift approaches $\frac{1}{2}$ radian.

Method	Output form	Max. gain	Comments
A-1 Delay & subtract	Frequency & sign	$\frac{1}{\omega}$	Delays of more than a few millisecs are costly
A-2 Count up and down	Number	$\frac{0.2}{\omega^2}$	$\omega < .002\ f_o$
B-1 Frequency tracking error	Frequency & sign	$\frac{0.3}{\omega}$	
B-2 Period tracking error	Frequency & sign	$\frac{f_b - f_o}{\omega f_o}$	$f_b > f_o$ and $f(t)$ small enough to track by this method. Say $f < 1$kHz.

A-1 is difficult to implement. B-2 is only suitable for certain frequencies. A-2 is the traditional method. B-1 is superficially unattractive, however it is quite often the case that a frequency tracking circuit is already present in the system for other purposes, and $f'(t)$ is then available at no extra cost.

If $f(t)$ exhibits jitter, then its derivative will be very noisy. Under such circumstances the extra smoothing of method A-2 is useful.

3.9 Non-linear Function Generation

3.9.1 Categories of function generator

It is not always easy to delineate those techniques (pulse-frequency) that lie within the scope of this thesis from those (e.g. pulse-number) that do not. In the area of non-linear function generation the distinction between pulse-frequency and pulse-number operations is particularly fine. Certain pulse-number techniques result (as a side effect) in the generation of pulse-frequency modulated signals. A distinction of practical importance is between essentially continuous processes (pulse-frequency or incremental) and 'one shot' processes (number and pulse-number) which have an identifiable beginning and end. Sampled-data processes link these two types.

Five categories of non-linear function generation may usefully be distinguished here, viz.:

(a) those where a number is generated from a pulse-number

(b) those where a frequency is generated from a coded number

(c) those where a frequency is generated from a pulse-number

(d) those where a number is generated from a frequency

(e) those where one frequency is generated from another.

The first category is amply described in the literature[302-306], and has acted as an historical parent to the other categories. In general, functions of a number may alternatively be functions of time elapsed, for pulse-number from a reference oscillator represents time elapsed.

Categories (b) to (e) will be considered below.

3.9.2 Frequency functions of a coded number

Linear number-to-frequency conversion was treated in section 3.4 of this thesis. Two non-linear conversions are fairly readily performed: polynomial and reciprocal functions.

Polynomial functions of number are obtained by first cascading frequency scalers to obtain the sequence of pulse-trains with frequencies $f_o = f_r$, $f_1 = \frac{y.f_r}{M}$, $f_2 = \frac{y^2}{M^2}.f_r$ etc. where y is the (coded) number, M is the scaler capacity, and f_r is the frequency of a clock train. Further scaling by constants followed by the merging of pulse-trains gives an output frequency.

$$f_x = f_r\left(a_o + \frac{a_1}{M}.y + \frac{a_2}{M^2}.y^2 + \ldots.\right) \qquad (3\text{-}106)$$

As the merging pulse-trains are all derived from the same clock train, they may be merged without using anti-coincidence gates, providing suitable small delays are introduced. The method only allows of fractional values for the scaling coefficients a_o, a_1, a_2 etc. Winston[344] used this approach to generate the time function $f_x = f_r(1 + at)$.

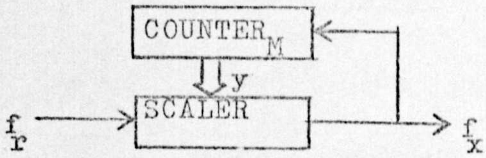
Reciprocal functions of number mean that output pulse-period is proportional to some number y. This relationship arises directly from the use of a frequency divider as described earlier (equation 3-36 and reference [330]). Every y^{th} clock pulse is output, so

$$f_x = f_r/y \qquad (3\text{-}107)$$

For good resolution, the clock rate f_r must be much higher than the desired output-frequency.

3.9.3 Frequency functions of pulse-number (or time).

Certain pulse-number to number operation can be operated with constant pulse-rate signals to give frequency functions of time. The circuit on the right gives an exponentially varying output frequency



$$f_x = f_r . y/M$$

If the counter counts up from an initial state y_o , then

$$y = y_o + \int_o^t f_x \, dt$$

giving

$$y = y_o e^{at} \quad \text{and} \quad f_x = ay_o e^{at} \qquad (3\text{-}108)$$

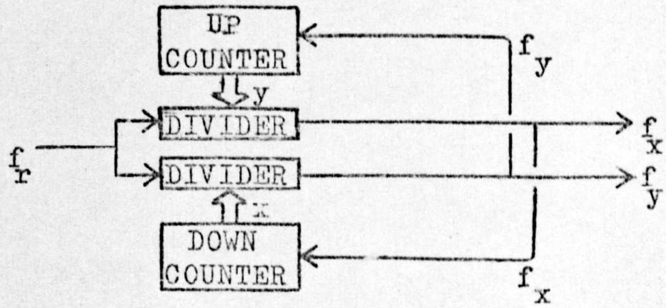
where

$$a = f_r /M$$

If the counter counts down from y_o , then

$$y = y_o e^{-at} \quad \text{and} \quad f_x = ay_o e^{-at} \qquad (3\text{-}109)$$

Martin has recently proposed[343] an exponential number-generator using dividers instead of multipliers.



So

$$f_x = f_r /y$$

and

$$f_y = f_r /x$$

The feedback count directions, and initial counter contents x_o and y_o give:

$$x = x_o - \int_o^t f_x \, dt \quad \text{and} \quad y = y_o + \int_o^t f_y \, dt$$

Solution of equations yields:

$$f_x = ax_o e^{-at} \qquad \text{and} \qquad f_y = ay_o e^{at} \tag{3-110}$$

where

$$a = f_r / x_o y_o$$

Both types of exponential generator described above are subject to rounding errors of varying severity, particularily when the numbers in the counters are small. Both types reach their limits when a counter overflows or empties; although repetitive exponential frequency sweeps can be arranged by suitable automatic resetting.

A circuit that has been widely used[303, 306, 345, 346] generates sinusoidally time-varying frequencies. The circuit has the same block diagram as that above, but with multipliers instead of dividers, so that

$$f_x = f_r . y/M \qquad \text{and} \qquad f_y = f_r . x/M$$

$$x = x_o - \int_o^t f_x \, dt \quad \text{and} \qquad y = y_o + \int_o^t f_y \, dt$$

then if $y_c = 0$

$$x = x_o \cos at \qquad \text{and} \qquad y = x_o \sin at \tag{3-111}$$

where

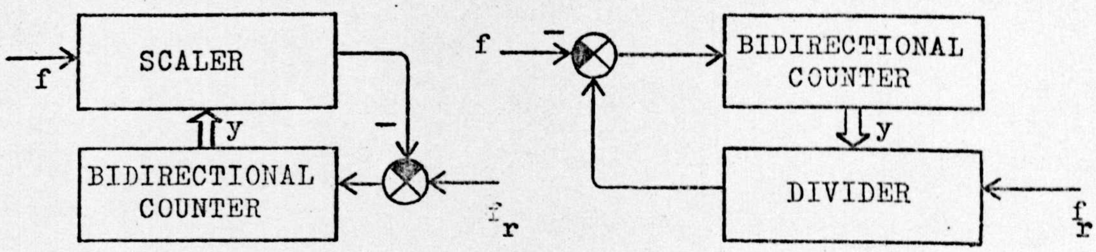
$$a = f_r / M.$$

By using bidirectional counters and subtracting a bias frequency of $\frac{1}{2}f_r$ from f_x and f_y, it is possible to make the configuration continuous running. However, used in this way, errors accumulate,[346] and the output frequency approximates to the form $e^{bt}\cos at$, where $b \ll a$[306]. It is therefore advisable not to subtract bias frequencies, but rather to reset both counters to their initial conditions the instant that the x counter empties. In this way a series of sinusoidal quadrants are described, which with suitable external switching will give complete sine or cosine waves.

In its basic form the configuration gives output frequencies of type: $f_x = a \sin at$ where $a = f_r/M$. Altering the clock rate will alter both the amplitude and time scale of the output, and where this is undesirable, additional scalers (fed with a constant clock rate) must be introduced.

3.9.4 Number functions of frequency

Various methods of linear frequency-to-number converters were described in section 3.7. Two of them involve a closed-loop frequency tracking arrangement. This could be modified to allow continuous period-to-number conversion, and two possible configurations are shown below:

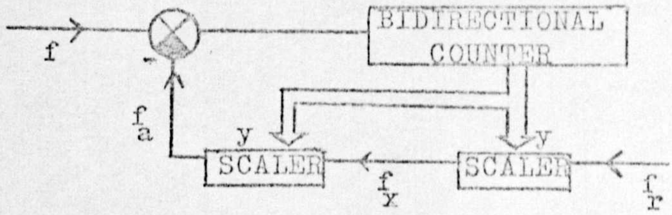


$y = M.f_r/f$ (at equilibrium) $y = f_r/f$ (at equilibrium)

In both cases suitable arrangements must be made to prevent the counter overflowing when f is very small. The left hand configuration will malfunction if $f_r > f$; the right hand configuration if $f_r < f$.

Other modifications of the tracking converters of section 3.7 permit other non-linear operations. In particular the polynomial frequency function generator of section 3.9.2 may be incorporated in a tracking feedback loop. A good example of this arrangement is the square-root generator shown over the page.

Provided f varies slowly it will be tracked by f_a which in turn equals $f_r . y^2/M^2$.



So near equilibrium

$$y \doteq K\sqrt{f} \qquad \text{and} \qquad f_x = \sqrt{f.f_r} \qquad (3\text{-}112)$$

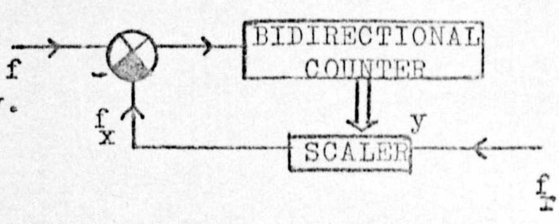
$$\text{where } K \doteq M.f_r^{-\frac{1}{2}}$$

If sampling type frequency-to-number converters are modified, more complex number functions of frequency can be obtained, Fixed period sampling converts pulse-frequency into pulse number, upon which a variety of non-linear operations can be performed.

3.9.5 Frequency functions of a frequency

Clearly frequency-to-number and number-to-frequency converters can be cascaded, and if one or both is non-linear, then the output frequency can be a non-linear function of the input one. Such an arrangement is likely to require extensive circuitry, and introduce serious delays and rounding errors. In some cases one or other converter produces the required output form as a side effect. The square-root frequency-to-number converter described by equation (3-112) yields not only the number y, but also a pulse train f_x whose rate varies with y.

The basic frequency tracking ciruit can be used to limit a frequency so that it does not exceed a reference frequency. As $f_x = f_r .y/M$, and M is



the capacity of the counter containing y, so f_x has a maximum value of $f_r(1 - 1/M)$. (The maximum value of y is M-1). Provided the counter is protected against overflow, f_x will closely follow f in value unless f_x meets its limit.

$$\begin{array}{llll} \text{If} & f < f_r & \text{then} & f_x \doteq f \\ \text{If} & f \geqslant f_r & \text{then} & f_x = f_r(1 - 1/M) \end{array} \qquad (3\text{-}113)$$

The circuit is subject to the usual tracking delays. While there are simpler ways of implementing equation (3-113), the one just described is attractive if the tracking loop has been introduced for other reasons such as frequency multiplication. The circuit shown in Fig. 3R(b) for instance, performs several functions including frequency-to-number conversion. In this circuit f_x is a somewhat delayed but jitter free version of the input frequency f, which is subject to both upper and lower limits.

$$f_x > f_b \quad \text{and} \quad f_x < f_b + f_r \qquad (3\text{-}114)$$

Unidirectional limiting in the absence of other operations is most simply performed using a frequency comparator whose output controls a switch. The switch connects either the input or reference pulse-trains to the output according to whichever is smaller. If jitter is likely to be present in the input pulse-train, the comparator will need several stages of pulse-canceller to distin--guish with a too-high mean pulse-rate and a too-high local pulse-rate caused by bunching. This sort of protection results in a phase hysteresis which causes delays in operation.

3.9.6 Rounding errors and delays.

Non-linear operations are generally more subject to rounding effects than linear operations. If rounding errors are expressed as a fraction of full scale (rather than of the instantaneous value represented by a signal), then rounding errors during linear operations are independent of signal value. Rounding error during non-linear operations will generally be some function of signal value, and be more severe at one end of the range of signal values than at the other.

Operations involving closed loops wherein a varying number controls a scaler are subject to a special category of error, known as a moving address error. This has been investigated by Cogman[346], Nightingale and Richards[339], who conclude that for small errors the counter addressing a scaler should count at a pulse rate much less than that which is being scaled. Of more significance than pattering error are malfunctions causing a permanent bias in an output. This can happen[346] when the input pulse-rate and the clock pulse-rate fed to an operational unit bear an integer or simple fraction ratio to each other. The resultant bias is related to a sort of error dead-band, which is generally fairly small unless the transition time through the (ripple) counter is comparable with the pulse period. The use of synchronous or semi-synchronous counters would be expected to reduce the likelihood of bias errors.

3.10 Frequency multiplication by factors greater than unity

Section 3.4 described digital scaling circuits for scale factors less than unity. Scaling a frequency up is a more complex operation than scaling down, because some form of prediction is usually required. There is no great difficulty in obtaining an output pulse-train whose mean frequency is some fixed multiple of that of an input pulse-train. If however the output pulses are to be fairly evenly spaced, simple techniques do not suffice. For example the feedbank arrangement described in 3.4.5[323] gives an output in which pairs of pulses may have a spacing independent of input frequency.

Before treating variable-frequency pulse trains, a group of signals will be considered whose frequencies can be more easily magnified.

3.10.1 Operations on signals with fixed frequencies or convenient waveforms

To generate a pulse-train whose rate is approximately a fixed multiple of that of a clock train is straightforward. An adjustable stable oscillator will serve. If the frequency ratio is to be maintained with zero mean error, some form of synchron--ising may be employed to 'lock' the oscillator to the clock train. Alternatively, each clock pulse can be used to trigger a number, say J, of output pulses spaced at intervals of $1/f_r J$, where f_r is clock rate. This latter technique may be extended to allow for integer variations in J, but it is in general rather clumsy and expensive.

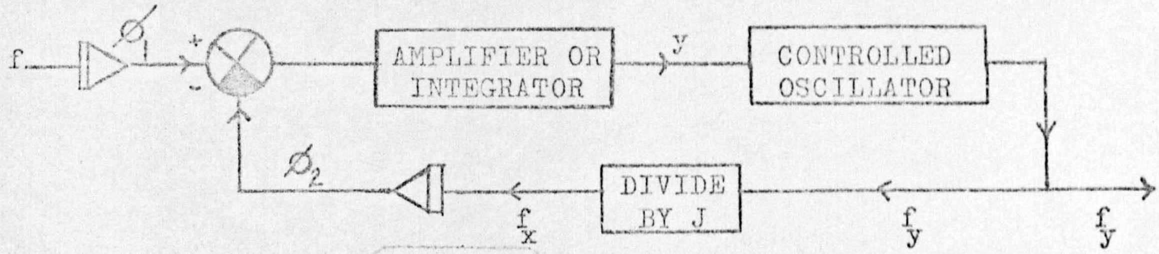
Margolin[348] discusses a number of non-linear analogue techniques for manipulating the frequencies of signals for which either waveform is fixed, or the frequency only varies by a limited amount (say ±20%). Pulse-trains, free from jitter and having pulse rates in a narrow band can be filtered to yield a fundamental sinusoidal frequency component. This can then be treated like other repetitive signals having fixed and convenient waveforms.

For instance a sinusoidal signal can be frequency doubled using a square law circuit, an analogue multiplier, or a phase sensitive detector also fed with a 90° delayed version of the sinusoid[347].

One cycle of a saw-tooth waveform can be dissected, using a set of amplitude detectors, into an integer number of stages, J. If the arrival at each stage is made to initiate an output pulse then J pulses will be output per cycle of the input saw-tooth. Suitable adjustment of detectors will enable the output pulses to be approximately equi-spaced.

3.10.2 Scaling up varying pulse-rates

To interpolate extra pulses uniformly into a pulse-train requires an estimate of the arrival time of the next incoming pulse. Some form of frequency-tracking is required. Most practical frequency-tracking systems, analogue or digital, take the form of a phase-locked loop. The phase difference between the incoming frequency and a locally generated one is used to modify the local oscillator. In the true phase-locked loop, the mean phase-difference is integrated over time, so that at equilibrium the phase-difference tends to zero. If a frequency divider is inserted in the feedback path of the loop, then a form of frequency multiplication is performed. (see over).



At equilibrium, $f_x = f$, $\phi_1 - \phi_2$ = constant, $f_y = J.f.$
The two frequency integrators may be replaced by a single one immediately before the controlled oscillator. Fig. 3R(b) shows such an arrangement (neglect the bias input f_b).

A true phase-locked loop may malfunction by losing lock (for example, a stepping motor malfunctions at high speed if not properly run up to that speed). The frequency tracking version that incorporates a phase amplifier instead of a phase integrator is not subject to this mode of failure. The loop is solely to amplify frequency, and not also to convert frequency to number. High loop gain in a digital frequency tracking loop (e.g. Fig. 3R(b)) means short counters and scaling circuits, and no need for pulse cancellers. Even though this allows some savings in equipment, frequency multiplication is still in equipment terms much more costly for factors greater than unity than it is for fractional factors. Moreover the dynamic response of the 'tracking' frequency multipliers just considered is always inferior to that of a fractional scaler.

CHAPTER 4

APPLICATIONS

Contents:

4.1 Frequency Modulating Transducers

4.1.1 Categories

There are large numbers of physical effects that might serve as the basis of transducers having frequency modulated or pulse-frequency modulated outputs. Relatively few such transducers have been developed, however, because frequency is not yet a standard signal form for measurements. A certain interest in frequency modulating transducers has existed for some years in the U.S.S.R., Hungary and Eastern Europe generally ... this has yielded a modest technical literature (e.g. refs. 401-5) and an attempt at standardisation[406]. In Europe and the United States design activities in the field have been extensive but disconnected, each frequency modulating transducer being a special purpose one. Some confusion is caused by the widespread custom of calling pulse-frequency modulating transducers 'digital transducers'. The term is inaccurate in this context.

The classification of FM and PFM transducers has been attempted by Knorring[402], while Agar[407] intended to give a complete review of developed types. Neither paper is very comprehensive, and a summary of transducers therefore seems appropriate here.

A preliminary distinction may be made between transducers in which frequency-modulation is intrinsic and those in which it is incidental. The latter category includes those electrical trans- which give a varying voltage or current that is subsequently converted to frequency. Were frequency a standard signal form, there would be no great difficulty in modifying existing measuring instruments to give an FM rather than analogue or coded digital output.

The mechanisms of intrinsically FM transducers can be grouped as follows

(i) resonance effects

(ii) integral-to-phase converters

(iii) random event generators

(iv) other mechanisms

4.1.2 Resonance

Many mechanical and some electrical systems exhibit resonances. These can be excited by the introduction of energy. All oscillations are accompanied by dissipation of work into heat, and only where this dissipation is slow can the oscillations form the basis of measurement. Lossy vibrations have a low Q (= quality) factor, and their frequency is easily modified by loading. Good transducers have low power consumption and are insensitive to loading.

The resonant frequency of a mechanical system is determined by the ratio of stiffness to inertia within it. Where one of these varies monotonically as the quantity to be measured, an FM transducer may be designed.

A taut wire is a simple distributed resonator. When plucked it vibrates in a number of modes and at frequencies that depend on modulus of elasticity and tension. Vibrations can be maintained electromagnetically, the maintaining circuit also serving to convert the vibrations to electrical form. For wires having constant Young's modulus, resonant frequency varies with the square root of tension. Taut wires have been used for the measurement of force, pressure, and extension[404,407]. The firm of Maihak (Hamburg) have developed a family of vibrating wire transducers for measuring strain, pressure, temperature and inclination inside civil engineering structures such as dams. The square-root characteristic can be used to compensate the square law characteristic of a venturi type flow meter (flow rate→differential pressure→frequency), and a patent exists for this usage with a lumped resonator[411]. Taut wires are not useful for measuring large deflections, as highly extensible materials have varying Young's modulus and are prone to creep.

Compressed struts have not been used in FM transducers, but would seem to offer possibilities for load cell design. The resonant frequency falls with load, reaching zero at onset of Euler collapse.

Complex resonances are observed in three dimensional bodies. Crystal oscillations have been traditionally used for time definition, the crystals being chosen and mounted to minimise the effects of pressure and temperature. A crystal can however be chosen whose resonant frequency varies appreciably, and linearly, with temperature[409]; one precision instrument is constructed on this basis[408].

A resonant system results if a current carrying coil is hung in a magnetic field. Galvanometers are usually designed so that oscillations are heavily damped. An alternative approach would be to omit the damping and exploit the oscillation frequency as a measure of the current-field product.

Gravity, particularly its variation, is conventionally measured by means of pendulum-like devices giving a frequency output.

Gas and liquid density meters are made[407] that contain mechanical resonances of variable frequency. A stiff tube containing the specimen fluid is vibrated. The frequency of resonance depends on the mass of tube and contents, and a calibration curve relates this frequency to the density of the specimen. A pulse-frequency calculator is used to approximate the calibration curve. The importance of a high Q factor is illustrated by the failure of these instruments with gas-liquid mixtures which absorb vibrational energy.

Complex vibrations occur in flowing fluids. Wave-frequency at sea is of slight interest in meteorology. A flow-rate transducer has been invented[410] that detects the longitudinal oscillations that occur in a flowing liquid downstream of a pipe constriction. Frequency varies linearly with flowrate, but there are practical difficulties in picking up the vibrations with microphones.

Vibrations of gaseous columns vary with so many gas properties such as temperature, pressure, gas content, that they are unsuitable for the measurement of any one property.

Frequency of resonance in electrical circuits varies inversely with the square root of capacity or inductance. These in turn may vary with temperature, size, spacing, and electric or magnetic material properties. There are several manual methods of accurately measuring capacity and inductance, and some of them are more convenient than interpreting variations in resonant frequency. However, where the measurement is part of an automatic control scheme, or where the primary transducer is remote from the observer, frequency of resonance is a convenient signal form. Commercial applications of variable capacitance transducers include the measurement of liquid level in tanks, of moisture content in paper and bricks, and of small displacements between moving parts. Inductive transducers are also used for displacement measurement.

4.1.3 Integral-to-phase converters

All frequency-modulating transducers give an output that is a cyclic function of the time integral of the input[402]. Only in certain FM transducers does the process of integration physically occur. Such devices might be classified as integral-to-phase converters, as the cyclic output may be assigned a conceptual phase. Tachometers belong to this class of transducer. The phase of the output (and hence the input position) is continuously observable with a.c. tachometers, but only intermittently observable with pulse tachometers.

To employ this type of transducer, the time integral of the variable to be measured must be readily available in a cyclic form. The table below lists some variables whose time integrals can be easily obtained, although not necessarily in cyclic form.

Entry	Variable	Time Integral	Analogous quantity
1	Speed	Distance	Angle
2	Angular speed	Angle	
3	Volume flowrate	Volume	Level
4	Mass flowrate	Mass	Weight
5	Force	Momentum	
6	Current	Charge	Angle, deposited mass, voltage
7	Power	Energy	

Entries 1, 2 and 3 correspond to commonly used transducers. The rate of passing distance markers is used as a speed measure. The markers may be mechanical (train control) or standing waves of coherent light (machine tool control). Wheels and propellers convert linear position into rotary angle, as in car and ship speedometers, allowing the use of tachometers. Volumetric flowmeters are made using turbines, tipping buckets and nominal pumps (positive displacement meters). Mass flowmeters are less common as mass flow can often be inferred from volume flow.

Entries 5, 6 and 7 do not find commercial expression. A design for a pressure transducer has been proposed [412] that is based on momentum balance. The change in momentum, due to pressure on one side of a freely supported diaphragm, could be balanced by a series of mechanical impulses on the other side. The proposed design entailed varying the size of the impulses until the diaphragm was stationary; the impulse frequency could have been varied instead. Feedback of diaphragm position could be used to alter the impulse frequency to obtain equilibrium.

Domestic electric meters commonly employ a motor (whose speed is proportional to current) and a tachometer. The tachometer pulses are totalised mechanically to indicate charge passed. Direct integration of current using a capacitor can be combined with arrangements to discharge the capacitor (and generate an output pulse) at a fixed voltage threshold. The integral pulse frequency modulators described in Chapter 2 are direct examples of integral-to-phase converters.

The quantisation of mechanical or electrical energy is too complex to justify integral-to-phase converters for power measurement. Power is usually inferred from indirect measurements, e.g. of torque and speed.

4.1.4 Random event generators

Many physical variables that are regarded as continuous are not so on an atomic scale. Radiation power and gas pressure are examples of statistical average variables. Normally it is not possible to detect the individual bombardments of a surface by quanta of energy or by molecules, and even were it possible the bombardment rate is not always a useful measure. Radiation power is quanta rate times quantum size. Pressure is bombardment rate times (twice) normal molecular momentum.

In nuclear physics, the event rates are often low enough to be countable, and very high gain power amplifiers have been developed to aid detection of the emission of single sub-atomic particles. Given a large population of assumed similar random event generators, the total event rate may be used as a measure of the individual event probabilities (actuarial records) or of the number of generators (nuclear physics). If the population is very large and homogenous, the overall event generation will be a Poisson process. The randomness of the events affects both transducer and instrument design. The transducer (e.g. radiation detector) should have a resolution time that is much less than the mean interval between events. An estimate of event rate is biassed by the factor: 1 - (resolution time) x (mean event rate). The instrument should allow for the variance in event rate by counting over a period much longer than the mean event interval, or by division of the event rate by an integer. A Poisson distributed pulse train has a pulse spacing whose standard deviation equals its mean. A pulse train obtained by division by k (i.e. every kth pulse of the Poisson train) has a more even spacing: standard deviation equals mean divided by $\sqrt{k}$.

Event-rate transducers are not strictly frequency-modulating ones, for they only detect events and thus change their physical dimensions.

4.1.5 Other frequency modulating transducers

There is a class of transducers that are superficially similar to the integral-to-phase converters of section 4.1.2. They possess however an inverting characteristic, so that output period is proportional to the measured variable. They may be called transit-time or charge-time transducers. In them a process is completed in a time that varies with the measured variable: every time the process is completed an output pulse is generated and the process is recommences. An example of this class is a 'sing-around' acoustic depth measurer, which on the assumption of constant sound velocity gives an output period proportional to fluid depth. Multivibrators or relaxation oscillators have a period proportional to electrical (or fluid) resistance or capacity: that is, a frequency proportional to conductance or chargeability. Attempts have been made to use the fairly straight conductance vs. temperature characteristic of thermistors as a basis for a frequency modulating transducers[413,414].

However the characteristic trimming required, the limited range over which thermistors are time stable, and the difficulty of manufacturing thermistors to close tolerances, have discouraged extensive developments.

The Döppler effect is quite widely used for velocity-to-frequency converters. A mixer or pulse-canceller is employed to subtract the transmitted signal from the received one: a frequency bias is introduced to permit distinction of positive and negative velocities. Modern asdics and radars are designed in this way to give both position and velocity information.

4.1.6 Experimental work on transducers

A number of frequency-modulating transducers were constructed in order to assess the ease of applying PFM techniques to process control. The intention was to show up unforeseen difficulties with the techniques rather than invent novel instruments. Consequently commercially developed components were used wherever possible. Five transducers were constructed to measure the key process engineering variables: temperature, pressure, flowrate, level, and valve-stem position. Deliberately, a variety of techniques was used, but the design criteria in each case were

(a) Output frequency range: 1kHz to 2kHz

(b) Output waveform: Steep sided pulses of amplitude 3 volts positive w.r.t. earth and of duration not less than 200 μsec

(c) Error (linearity and repeatability): $< \frac{1}{2}\%$ of range.

The main features of the five transducers are tabulated below:

Primary Variable	Range (Units)	Secondary Variables	Notes	Class of Transducer*	Notes
Temperature	0 → 100 (°C)	Length	Crystal oscillator	Resonance	Not linear
Pressure	0 → 400 (kN/m^2)	Strain. Resistance	Bridge as gain element	Charge-time	Drift in zero point
Flowrate	0 → 0.8 (l/sec)	Angular velocity	Tachometer	Integral-to-phase	Commercial transducer
Level	0 → 1 (m)	Angle	Shaft-encoder	Coded number to frequency	Gray code
Valve position		Resistance		Integral-to-phase	Coarse quantisation

* within the context of the preceding sub-sections

The design criteria were not easy to meet, nor was it possible to operate all the circuits at the low power levels associated with intrinsic safety certification. Since these instruments were designed (in 1967/8) the quality of electronic components has improved, particularly in respect of linearity of analogue circuits and low power consumption in digital circuits.

Obtaining a suitable frequency range can be difficult. The range 1kHz to 2kHz was chosen for the following reasons: having no d.c. component, power can be sent to a transducer over the same lines as signal is returned from the transducer; much electrical 'noise' is concentrated at low frequencies such as 50 Hz; 2kHz can be transmitted over almost any cable (including telephone lines) without attenuation or reflection; 1kHz span allows resolution of 0.1% full scale in a 1 second count. Adding in the zero frequency of 1kHz to integral-to-phase converters such as the turbine flowmeter has to be done after conversion - giving a jittery output train. Addition of bias prior to conversion is usually impractical: a bias flow cannot easily be injected into a flowmeter! The use of a frequency offset may exacerbate the problem of 'zero drift'. When the frequency offset is n times the frequency span, a 1% drift in the former represents an n % drift in the zero point of the interpreted measurement.

As a result of the design calculations performed for these five instruments, it was concluded that the cost of processing an electrical variable such as resistance or voltage into a biassed frequency could be achieved at a component cost of £5 (1971 prices). Such processing would not increase the errors due to the primary sensing element by more than ½%. Number-to-frequency conversion at 8 or 9 bit resolution would cost about £10 in components ... integrated circuits for this purpose have appeared on the market (Texas, Plessey) during the last year.

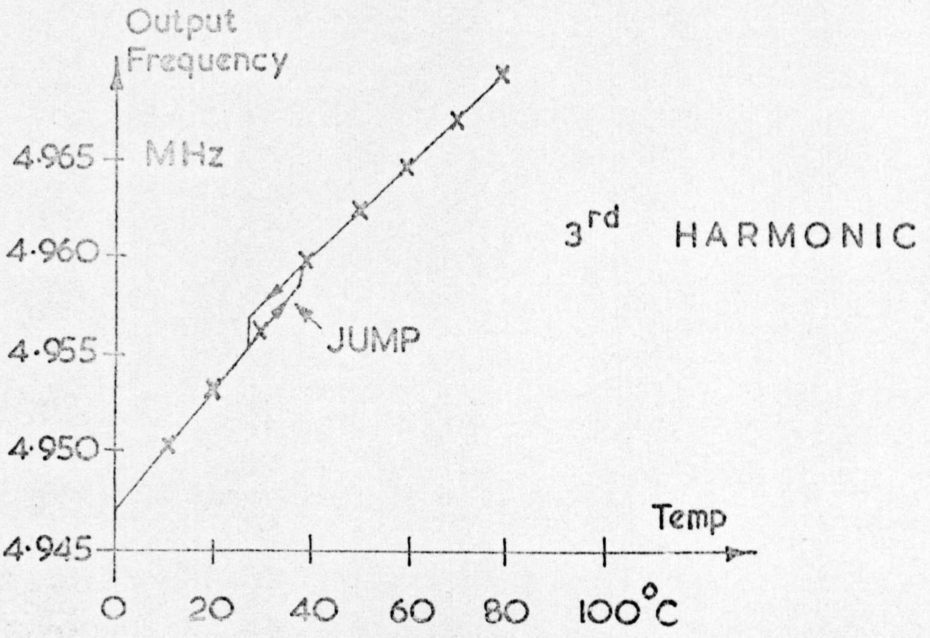
4.1.7 The temperature transducer

Temperature is the most commonly measured process variable, and a very wide range of physical phenomena are employed in temperature transducers. Two phenomena that are suitable for reasonably simple frequency-modulating transducers are change of resistance with temperature and change of length.

Resistance thermometry is well developed, both with respect to the materials with small linear positive temperature coefficients (e.g. platinum), and to those with large non-linear negative temperature coefficients (i.e. thermostors). So temperature → resistance → voltage → frequency conversion would involve a sequence of known techniques. This sequence seemed rather long, and therefore susceptible to cumulative errors, and is similar to the sequence under investigation for pressure measurement. Attention was therefore given to the shorter sequence temperature → strain → resonant frequency.

Quartz crystal thermometry has been known for some years, but has only been developed commercially for very high precision applications.[408] It was hoped that accuracies adequate for process control (say $\frac{1}{2}$°C) could be obtained using very simple circuits, and fairly cheap crystals. This hope was not fulfilled.

A crystal was obtained, cut as specified by Smith and Spencer[409], and operated as they recommended in the third harmonic mode. A significant jump discontinuity and change in slope was discovered in the frequency-temperature characteristic. This is shown in the upper graph of Fig. 4A. Operated at the fundamental frequency, the discontinuity disappeared, but the linearity was not within 1% over the interval 0 to 100°C, see lower graph of Fig. 4A. Temperature was found to affect not only the resonant frequency, but also the 'activity' of a crystal. To maintain the oscillator on the verge of oscillation, and



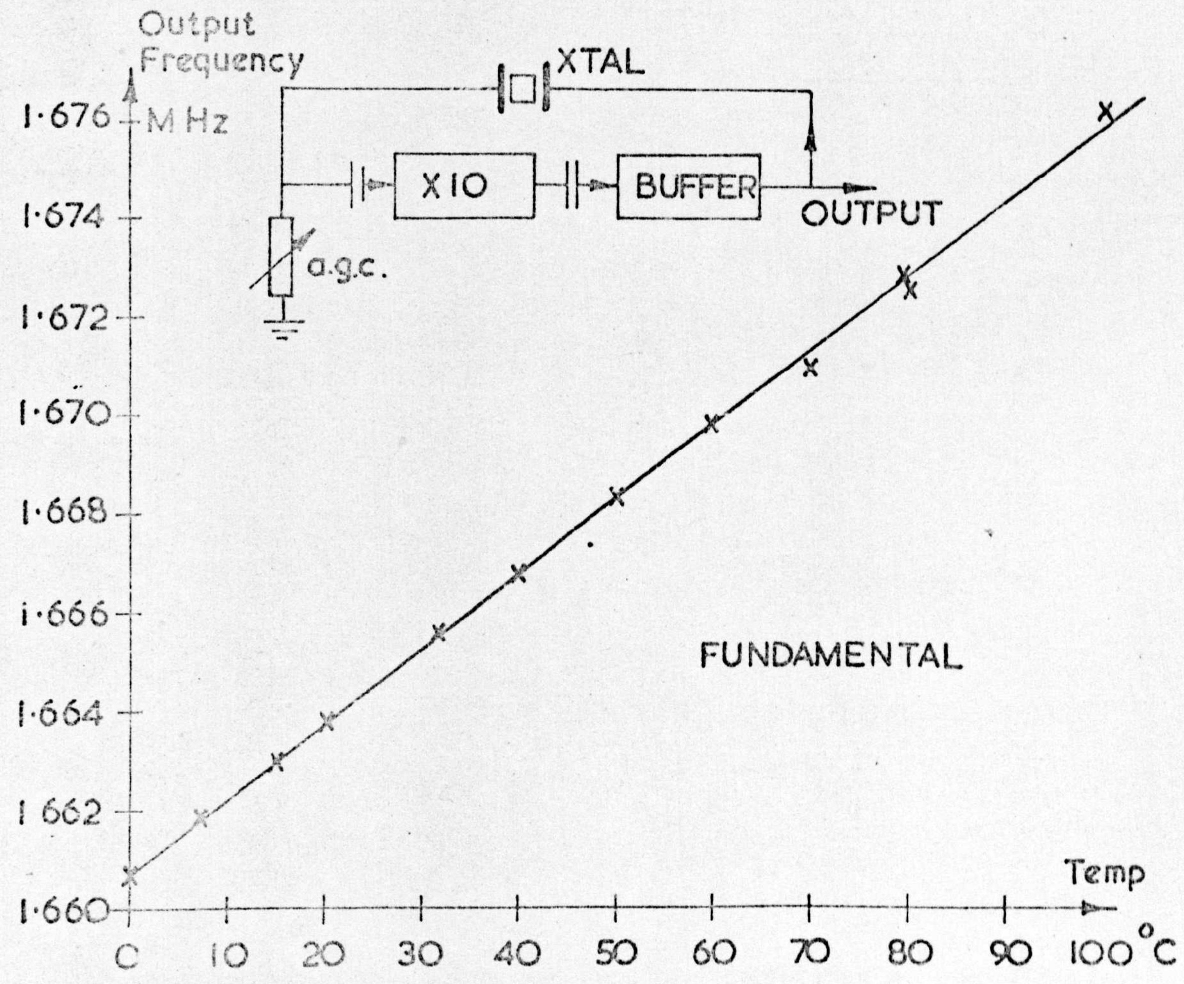


FIG 4A CRYSTAL TEMPERATURE TRANSDUCER – CALIBRATIONS hence at the peak of resonance, an automatic gain control is necessary to compensate for changes in crystal activity. Absence of this gain control causes frequency biases corresponding to errors of up to $\frac{1}{2}^{\circ}$C.

Having obtained a frequency that varies slightly with temperature, it is then necessary to subtract a stable reference frequency to obtain an output in an appropriate frequency band. The reference oscillator crystal would not normally be subject to the same or as large temperature range as the measurement crystal. Even so, the temperature coefficient of the reference oscillation must be much below that of the primary oscillation. The primary crystal used was found to have a temperature coefficient of about 100 parts per million per $^{\circ}$C, requiring (for $\frac{1}{2}^{\circ}$ accuracy) that the reference frequency should not vary by more than 50 ppm over the ambient temperature range. This condition is not an onerous one for an ambient range of 50°C, so that the expense of an oven is spared.

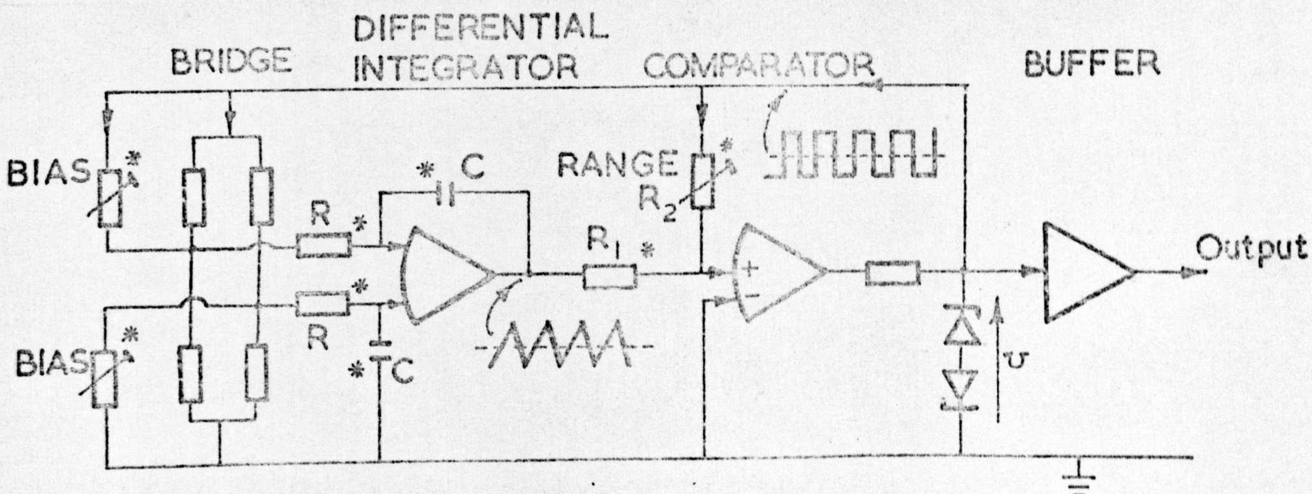
Crystals are usually mounted in a vacuum to avoid the influence of pressure variations or gas density upon frequency. The crystal used for experiments was so mounted and consequently has an inconveniently high thermal time-constant of about 3 minutes. The precision crystals used in the Hewlett Packard instrument[408] are mounted in an inert gas within a very rigid shell, their time constant is about $\frac{1}{2}$ minute.

In conclusion the production of a low-cost, full encapsulated crystal-oscillator temperature transducer with adequate performance seems rather distant. The frequency sensitivity is high enough, but crystal preparation and mounting require special care. Some of the oscillator circuitry would need to be apart from the crystal if temperatures over 100°C were contemplated, yet the cable linking crystal to oscillator must have a closely defined capacitance. The recent appearance of r.f. amplifiers with a.g.c. in integrated circuit form is helpful to the development of a compact transducer. The power consumption of a crystal oscillator can be made very low and there is no difficulty in operating over a single pair of wires through an intrinsic safety barrier.

4.1.8 The pressure transducer

The most common pressure transducers having electrical outputs are of the force-balance type. They are not very linear and their power consumption is high. There are also a number of pressure sensing elements involving the conversions force→displacement→variation in resistance. Of these, designs using strain-gauges are more linear than those using large displacement bellows or Bourdon tubes. The temperature dependence of strain gauges can be compensated by the use of bridge networks.

A strain-gauge bridge pressure transducer (Ether Ltd., Type UP.4) was incorporated into the feedback of a square-wave oscillator. The circuit diagram is shown below.



Bridge-gain to frequency converter ($\upsilon = +V_1$ or $-V_2$)

The frequency of oscillation is determined in the absence of bias by:

the bridge 'gain', g,

the time constant, $T = CR$,

the ratio, $r = R_2/R_1$

the ratio, $b = V_2/V_1$ (where v is clamped to $+V_1$ or $-V_2$)

and

$$\text{frequency} = \frac{rg}{T(2 + b + \frac{1}{b})} = f \tag{4-1}$$

As $df/db = 0$ when $b = 1$, the zener diodes shown in the circuit were matched so that $V_2 \doteq V_1$, provided $0.9 < b < 1.1$

$$f = \frac{rg}{4T} \pm 0.3\% \tag{4-2}$$

There is no difficulty in selecting and operating similar zeners to maintain b within this range.

The frequency of oscillation is susceptible to variation through several causes, the principal of these being:

- variation with temperature of any of the 8 components marked with an asterisk in the sketch above,
- drift in the comparator,
- drift in the differential integrator,
- substantial changes in bridge resistance (which is assumed much less than R).

The linearity of the strain-frequency characteristic depends on

- the linearity and bonding of the strain guages,
- the common mode rejection, input impedence and gain of the differential integrator amplifier.

The circuit was constructed using resistors and capacitors of mean temperature coefficient 25 ppm/°C and a type 709 integrated circuit operational amplifier for the integrator. The temperature coefficient of frequency was measured as about 100 ppm/^{o}C. The overall pressure-frequency characteristic was found to be linear within ½% over the required range. However over three years the zero position (1 kHz) was found to have shifted by 2%, of which only 0.1% is attributable to the strain guage.

The use of a bridge as primary sensor has required a relatively large number of stable components. The constraints upon the operational amplifier are not difficult to meet with contemporary integrated circuits (100db common mode rejection, gain $> 10^4$, current offset $\leq$ 10nA). The requirement for stability in passive components would be expensively severe if 0.1% accuracy were contemplated.

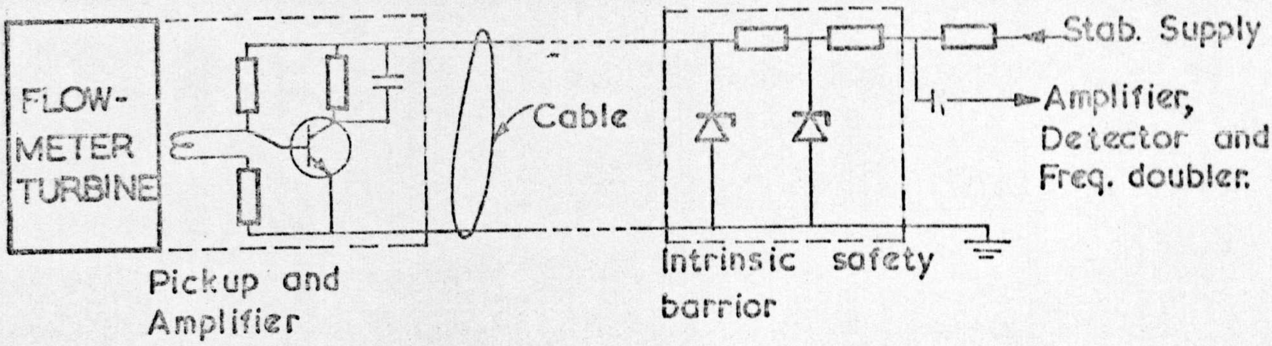
The transducer as made requires a 30 volt supply which is smoothed and divided to drive $\pm$ 12 volt operational amplifiers. This potential is too high for 'intrinsic safety' operation.

<u>4.1.9 The flowmeter</u>

A commercial turbine flowmeter (EFM type B/⅝/15) was augmented by digital circuitry to give a particular output range (1kHz to 2kHz ≡ 0 to 0.8 l/sec). An auxiliary 1kHz oscillator was employed for the zero bias. The rescaling factor for range was 1.425. This was implemented by first doubling the frequency of the turbine pulses and then multiplying by the binary factor 0.1011011. The former operation required analogue circuitry, the latter a binary rate multiplier. As it is possible to manufacture turbine flowmeters to any particular scaling, the rescaling exercise was only of local interest.

Turbine flowmeters are accurate within ½% over the interval 3%-100% of design range; biassing and rescaling in the pulse-frequency domain does not alter the accuracy.

The flowmeter included a single stage amplifier to aid the recognition of the turbine pulses at low flowrates. The amplifier was supplied with power via an intrinsic safety barrier, and the detected pulses returned over the same line. This method of working and the use of a suitable amplifier at the receiving end proved satisfactory for the measurement of flows as low as 1% of full range.



4.1.10 The liquid-level transducer

A small liquid level transducer was constructed using a stabilised float, connected by a taut wire to a drum. The range of motion was 1 meter but could without difficulty have been extended. The wire was tensioned by a 'constant tension' spring; the slight variations in wire tension were calculated not to alter the float's immersion by more than 1 mm. Similarly the expected variation in water density (4%) due to temperature changes was calculated not to introduce errors as great as 1 mm. The float's area-to-immersed volume ratio was made high. The float was guided vertically to avoid cosine errors caused by lateral displacement. The expansion temperature coefficient of the suspension wire was adequately low.

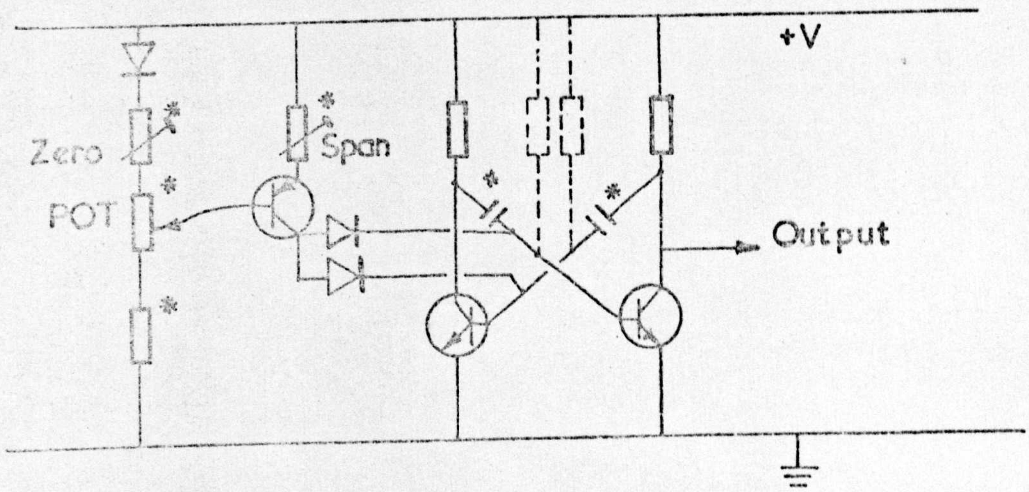
The wire was wrapped round an accurately machined drum to which a shaft encoder was attached. A nine-bit gray coded number was transferred in parallel to a gray-to-binary code converter and then to a binary rate multiplier. In this application code edge uncertainties are of slightly less importance than they would be in a digital sampling system, due to the continuous nature of the frequency scaling operation. Integrated circuit gray coded rate multipliers have very recently been advertised; they would fit this application.

The depth-to-binary conversion was found to be accurate to 0.3% including quantisation. This statistic was however obtained prior to fitting the float due to the difficulties in calibration within a tank.

4.1.11 The valve position transducer

The motorised valve used in the experimental rig was supplied with a potentiometric feedback element. Unfortunately this was found to be wire wound, having only about 50 turns over the valve range of interest (0-60% open). There was consequently no possibility of obtaining ½% accuracy.

As valve position is not a variable that needs to be known very accurately - it occurs in the forward path of a control loop - the valve was not modified to accommodate a better feedback potentiometer. Frequency encoding was effected via a temperature compensated multivibrator:



All the components marked with an asterisk directly affect the output frequency and were chosen to have a low temperature coefficient. The general arrangement reduces the influence of supply line potential variations, $\Delta f/f \doteq 0.1 \Delta V/V$, but the measured temperature coefficient of output frequency was high, about 1000 ppm/$^{\circ}$C. Zero drift over 3 years was 0.3%.

4.1.12 Conclusions concerning transducers

The target of ½% accuracy was not a particularly high one, yet the combined effects of quantisation, non-linearity, component ageing and variation in ambient temperature made it difficult to attain. Of the five transducers constructed, only those measuring flowrate and level attained and are believed to have maintained the desired accuracy. These two transducers owe their accuracy to mechanical rather than electrical components.

The transducers employing essentially electrical conversions to frequency were found to require relatively large numbers of stable passive components, as well as high gain amplifiers. Some means of limiting ambient temperature variations seems desirable, perhaps a low power oven to convert an ambient range of 0°C-50°C to say 25°C-50°C. The use of such an oven is practical only if the circuitry is very compact, for example a hybrid integrated circuit.

An alternative approach for electrical conversions would be to develop a single well-compensated converter and match any electrical sensing element to it. Probably the most suitable converter would be of the voltage-to-frequency type, accepting an input in a standard range and generating a frequency in a standard range.

Another useful 'standard' circuit for incorporation in transducers would be a frequency inverter, performing the operation $f_{out} = (f_{ref}/f_{in}) + f_{bias}$. This would complement a number of fairly simple devices of the 'transit time' type, for the measurement of distance, capacitance etc.

Digital inversion circuits were described in Chapter 3, they are somewhat complex and consequently expensive.

Present day process instrumentation has an accuracy of the order of 1%, varying with application and the variable being measured. The five transducers constructed had accuracies of this order, so that with proper engineering development they should all have been capable of ½% performance. No strong conclusions can be drawn from their performance, but it did indicate that there are no major obstacles to process instrumentation using an f.m. standard signal.

Intrinsic safety has been mentioned several times. The trend towards compactness and lower power consumption has brought many electrical transducers within range of being intrinsically safe. Unfortunately there is not yet international agreement over the meaning of this term[415,416], which is variously interpreted in different countries and different industries. Generally it implies an absence of large energy storage devices, electrical potentials nowhere exceeding particular values (e.g. 10 volts), and a low power consumption (e.g. ¼ watt). Every certified device in a dangerous zone (e.g. presence of hydrogen gas), that communicates with apparatus in an unrestricted zone, must do so via a barrier that will block excessive power or potentials passing from the latter to the former. Each non-earthed line requires a separate barrier. Suitably certified barriers are not cheap (say £15), and can distort signals that pass through them. 'Live zero' frequency modulating transducers are particularly suitable for operation via such barriers, because only one pair of wires is needed for each transducer and the signal will bear some distortion.

4.2 Instrumentation

4.2.1 The use of PFM in instrumentation

Pulse-frequency modulating transducers have been described at some length in the previous section. Certain of these transducers are sufficiently superior to other types to justify their immediate choice. The majority of FM and PFM transducers are not so markedly superior, and are only employed when the total instrumentation context favours them. Instrumentation may be divided into transduction, telemetry, computation, recording and display.

The general communication properties of PFM were examined in Chapter 2. A property of importance in telemetry is that of transmitting very low frequency signal components without distortion - PFM transmission possesses this property. Any binary or pulse transmission scheme can be designed to give lower error rates than direct transmission. For accuracies greater than $1:10^5$ serial or parallel binary encoding is required, while accuracies less than $1:10^2$ are easily handled by direct or A.M. transmission. FM and PFM suit the range $1:10^2$ to $1:10^5$, or lower accuracies over very poor channels. The complexity of PFM encoding is usually much less than that of serial binary encoding. PFM signals are not susceptible to corruption by passage through a simple multiplexor (in the way that analogue signals are).

Computation is a feature of much instrumentation: common operations being rescaling, function generation, integration, averaging and multiplication of one measurement by another. Computation with PFM signals, like telemetry, is generally more accurate than computation with analogue signals and less complex than computation with coded binary signals.

With respect to display and storage, PFM signals are usually converted to coded binary or decimal form. The advantages of digital display and the convenience of digital storage have been widely recognised by instrument manufacturers in recent years. The problem of radix conversion (e.g. base 2 to base 10), which complicates some fully digital instrumentation, is largely avoided with PFM. A pulse-frequency can be as easily converted to a decimal number as to a binary one. Where a single display has to convey rate as well as value information, analogue displays are preferred. The usefulness of PFM in display lies in its central position, i.e. its ready convertibility to all other forms.

4.2.2 Examples of PFM instrumentation

In section 4.1 a number of applications of PF modulating transducers were discussed. The following examples have been chosen to illustrate the use of PFM in instrumentation where other considerations than transducer design are the critical ones.

(a) Measurements in moving parts. The measurement of the temperature of a piston, or of the torque in a rotating shaft, cannot normally be communicated to an observer directly through wires. Either some form of slip-ring or electromagnetic induction has to be used. The transducer has to be both light and reliable, while the transmission may be in an electrically noise environment. Frequency modulating transducer-transmitters are employed, powered by small batteries. Many extensions of this technique to medicine and zoology are described in a book by Mackay that has been already cited[219]. Coded digital transmission is not used because the transmitter would be too bulky, expensive, unreliable and power-consuming.

(b) Domestic metering. The measurement of consumption (over a period) of gas, electricity and water usually employs pulse-frequency transducers and mechanical integrators (totalisers). The meters have to be cheap, robust and reliable and yet achieve accuracies for which analogue techniques are barely suitable. The mechanical pulse counters are easily read and comprise non-volatile storage. For administrative convenience, resolution greatly exceeds cumulative accuracy. The pulse-rates used are very low and unsuited to transmission.

(c) Resonance scanning. Mention has already been made of a range of vibrating wire transducers[411] for use in civil engineering structures. Measurements of strains, temperatures and inclinations in structures may continue over several years, so that transducers cannot be self powered. The passage of time causes the decay of low cost wiring until communication with a transducer may be over a highly resistant, leaky and even intermittently broken path. A transducer's resonance can be observed as a change of impedence during frequency sweeping. The emphasis in this application is on reliability and operational simplicity at low cost; accuracies are generally low.

(d) Rescaling. Machine tools are provided with high accuracy analogue position meters in the form of calibrated dials and verniers. Numerically controlled machine tools require electronic position transducers, and both incremental and absolute encoders are employed. A number of conversions may be required in the instrumentation, such as binary-to-decimal conversion or imperial to metric conversion. Work is proceeding in the design of an operator's display to be driven from binary shaft-position encoders. It will allow the display of a position as an absolute or relative decimal quantity, in imperial or metric units. Similar displays are already marketed; but using parallel binary arithmetic they are complex and expensive. Conversion of encoder output into frequency permits simple circuitry to be used for the display and scaling operations.

(e) Variable multiplication. The direct measurement of power, or of mass-flowrate is not normally possible, and these quantities must be inferred from measurements of current and potential (or speed and torque), and of volumetric flowrate and density. There are a number of other common measurements that are obtained by multiplication or division. The multiplication of two pulse-frequencies was shown in Chapter 3 to normally require conversion of one pulse-frequency into a number. Schemes for both multiplication and division were indicated. Where both pulse-trains comprise narrow pulses, and where one or both is sufficiently random, a logical 'OR' gate suffices to perform the multiplication. Torque is fairly readily transduced into a frequency; speed, volume-flowrate and fluid density are regularly so treated. An English instrument multiplies together the pulse rates from a turbine-flowmeter and (after rescaling) a density meter, to give mass flowrate.

(f) Alarm scanning. Smiths Industries market a limited range of PFM transducers (and converters) for process instrumentation. The range includes adjustable frequency-sensitive switches which are intended to raise an alarm when a transducer frequency goes outside prescribed limits.

(g) Environmental measurements. Measurements of the natural environment are characterised by low bandwidths and by geographical dispersion. The G.P.O. has recently developed a new telemetry service (Datel 400) at the instigation of the National Environment Research Council. Remote transducers are designed to emit a tone (related to the measurement) when they are interrogated by telephone. As each interrogation need not last more than (say) 10 seconds, there is no tariff advantage in using serial binary coding and frequency-shift keying to further reduce the time for data transfer. Encoding into tone is more cheaply done than serial binary encoding.

4.2.3 Computers in instrumentation

Computers, as their cost has fallen, increasingly form part of instrumentation systems. They are used to compensate known transducer imperfections, to relate different measurements and organise efficient information displays, to test measurements as they are taken (e.g. for unwelcome values), to extract long-term statistics and to maintain sophisticated memories. In many situations, the detail with which measurements are recorded should be some reciprocal function of their age. A computer can be programmed to maintain a set of memories (representing for example 'last 1 minute', 'last 30 minutes', 'last 24 hours' and 'last 30 days'), each one a set of statistics derived from the previous one.

Computers are also being used in a more active role to periodically revise measurement calibrations. Thus in a runway visual-range indicator[417], where a lamp and a photocell are used to measure loss of visibility, the photocell is periodically subject to full, zero and 10% light intensities. The regular measurements are briefly interrupted while the photocell calibration table is thus recomputed. Similarly, the pressure transducer that was constructed (as described earlier), was connected to both a pressure vessel and to atmosphere through separate solenoid valves. A computer programme was written to regularly record the pressure in the vessel, and also to periodically switch over the solenoid valves and measure the transducer output at atmospheric pressure. Any change in this zero point was caused to modify the equation used in the computer to translate transducer output into pressure. Not all transducers lend themselves to this technique, which is particularly relevant to (chemical) analytic instruments.

Where a computer is used with PF modulating transducers there are several ways of entering the outputs of the latter into the former. A digital computer is not a continuous but a sampling element in an instrumentation system, so that pulse counting would seem the most appropriate form of frequency-to-number conversion. Pulse-counting has good jitter averaging properties. Pulses can be accumulated in a computer word by the use of an interrupt feature, and for a small number of transducers this method has a few attractions. For large numbers of transducers or high pulse rates the method makes unacceptable demands upon central processor time.

In computer installations that are primarily geared to handling analogue signals (via a multiplexer and analogue-to-digital converter), the isolated PFM signals is sometimes first converted to analogue form. This is inefficient, but is done to remove the complexity of an extra category of signal.

The most efficient way of entering data to a computer is as serial or parallel binary coded words. In the same way as analogue-to-digital conversion is performed in a peripheral unit, frequency-to-digital conversion should be done without involving the central processor. Because F→D conversion is much simpler than A→D conversion, it should not be necessary to multiplex all inputs to a single converter, as is analogue practice. At the instant of transfer into the computer, a digital word should be unchanging. This is easy to arrange, particularly when the word is being taken out of a synchronous or semi-synchronous counter. There seems little need for the Gray coded counters that have at various times been proposed for this application, while there is a tangible time saving if subsequent Gray-code to binary conversion is obviated.

Two transfer techniques may be used. In the first, the incoming pulse train feeds a synchronous up-counter, which is periodically frozen, copied into the computer store, and reset to zero. In the second the incoming pulse train feeds a tracking frequency-to-number converter; the number may be read into the computer at any time provided the tracking loop is momentarily frozen. The first technique gives much simpler circuits, viz. an up-counter instead of a bidirectional counter, a scaler, a clock and a coincidence gate. However the number transferred is the product of the measurement and the sampling time, requiring the programme scaling to be adjusted if the sampling rate is changed. The second technique allows any bias frequency to be removed, and enables the computer programmer to treat measurements exactly as if they were stored numbers. Both techniques can be modified to permit the switching of several pulse trains sequentially to a single counter or converter. If N signals were so multiplexed, the modulation index of each signal must be N times higher than would be necessary using separate counters or converters. The first technique is the more appropriate if multiplexing is employed.

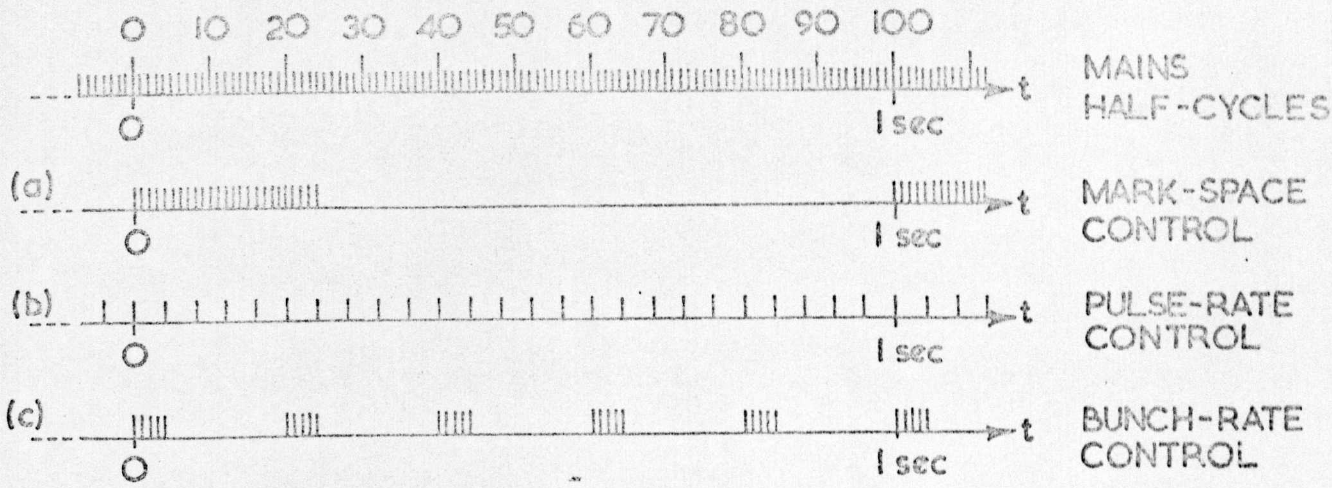
4.3 Controller design

4.3.1 Choice of actuator

An actuator is primarily a transducer, converting the dimensions of a signal from its standard communication form into those of the controlled variable. An actuator is often also an amplifier, insofar the power levels associated with the control signal are usually much less than those of the controlled variable. A PFM control signal is not particularly convenient, as the controlled variable is frequently a continuous or 'analogue' one.

Two actuators that directly transduce and amplify pulses or events are the stepping motor and the dosing pump. The former is a type of synchronous motor that, given appropriate drive circuits, can be made to advance by a fixed angular increment upon each command pulse. The latter will deliver a fixed fluid dose upon receipt of each command pulse. The stepping motor thus gives a velocity proportional to command pulse-rate, torque being assumed high enough to prevent loss of synchronism. It is used in certain machine tools and as a component in pulse-number to air-pressure converters. Where position is the controlled variable the stepping motor acts as an integrator of signed pulse-rate. The dosing pump similarly gives a flowrate proportional to pulse-rate but a total addition that is the integral of pulse-rate. The application of pulse-rate control to fuel-injection (via a dosing pump) or sparking is not a satisfactory form of speed control for internal combustion engines.

Alternating current power supplies may be regarded as a source of energy pulses supplied at (in Europe) 100 per second. The size of each pulse is a function of source potential difference, load resistance and sometimes load inductance. Modern practice with thyristors allows the selective passage or blockage of individual or groups of energy pulses. This is known as integer cycle control and generates far less electrical noise than the earlier phase control methods.



(a), (b) and (c) show different ways of selecting ·25% of the mains power to apply to a heater.

The power dissipated in the load will be a maximum when every energy pulse is passed by the thyristor switch. A power of N% of the maximum is usually implemented by mark-space methods: for example in one second N pulses are passed and 100-N are blocked. In simple mark-space control the total (mark + space) interval is chosen to be a fraction of the dominant time constant of the power dissipating device, such as heater or d.c. machine.

An alternative approach is to fix the mark period as one energy pulse-period (= one mains half-cycle time) and to vary the space time in units of energy pulse-periods. This is the same as making the mark half-cycles correspond to a synchronised pulse-train whose mean frequency determines the mean power. 100 Hz corresponds to full power: 25 Hz corresponds to 25% power and is shown in line (b) of the sketch above. Here then is a mechanism for converting a pulse-rate of not more than 100 Hz into a proportional electrical power. It works by turning the power on for any given half-cycle if a command pulse was received in the previous half-cycle.

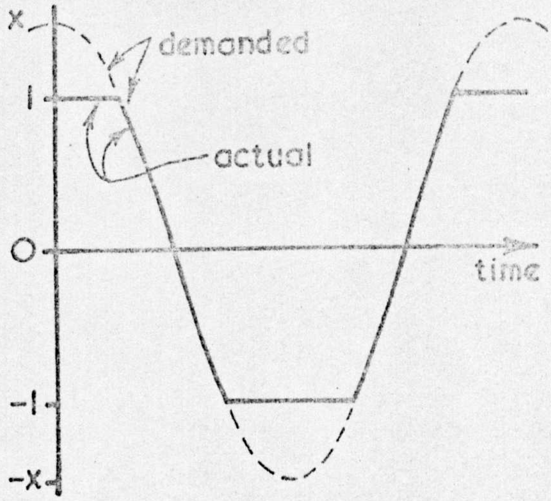
There are sometimes advantages in not introducing so many transient conditions as result from switching individual mains half-cycles. In such cases the control pulse-rate can be divided by some integer M and each divided pulse used to initiate a bunch of M mains half-cycles. This is shown in line (c) of the sketch, for 25% power and M=5. Over most of the power range this bunch-rate control has properties intermediate to the other forms. At very low power (< M%) it may result in significant output ripple.

It has not been established whether pulse-rate power control has been used in commercial equipment.

The commonest form of process actuator is the flow valve, either pneumatically or electrically moved. Pneumatic valves are more common, especially in hazardous areas; the relative merits of the two types are very dependent on application and availability of supplies. Using a stepping motor converter the valve position will be approximately the integral of the signed control signal's pulse-rate. Exceptions arise if the valve should hit its limits of motion. Some thought has been given to designing a pneumatic positioner that would receive its command signal as a pneumatic pulse-frequency, and its power supply from a steady pressure air-line. Unfortunately the very low bandwidth of long air lines (a few Hertz) seems to rule out this approach.

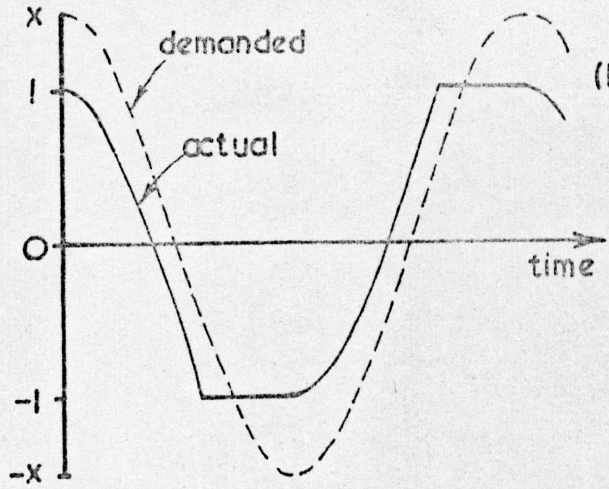
Operation of an electric actuator requires a three level signal corresponding to the valve states of opening, stationary and closing. These signals correspond roughly to the modulus of the time derivative of desired valve-position. The taking of a modulus involves loss of information, and in general the stepping motor arrangement is preferable to a ternary arrangement. The problems of ternary control are discussed in the next section.

Incremental actuators exhibit interesting limit behaviour. If a motor driven actuator reaches an end stop, it necessarily must disregard further commands to increment its position in the same direction. For example the command pulses to a stepping motor are ignored or 'spilled' once the limit of motion is reached. If the command pulses are reversed in sign, however, the motor responds to them immediately. The relationship between the commanded motor position and the actual position is illustrated in Fig. 4B (b). The way the actual device position pulls away from the limit at +1. when the <u>slope</u> of the demanded position changes sign, results in a degree of phase advance. Fig. 4B (a) shows normal amplitude limiting, while Fig. 4B (c) represents both conventional and



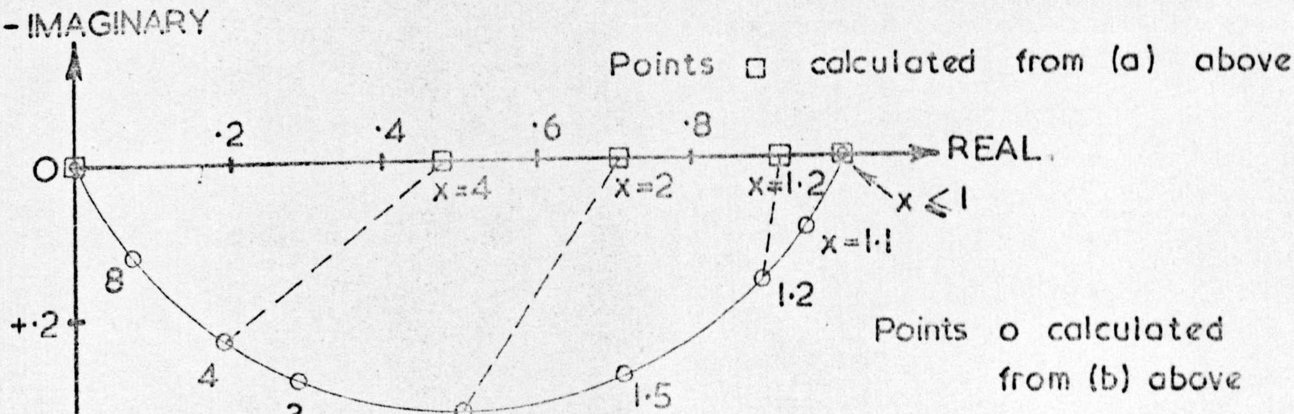
(a) Response of a conventionally limited (saturating) device to a steady sinusoidal demand of amplitude x

(x > limit)



(b) Response of a limited device that 'spills' any excess demand, to a steady sinusoidal demand of amplitude x

Note the 'phase' advance.



(c) Complex describing function for two sorts of limiter

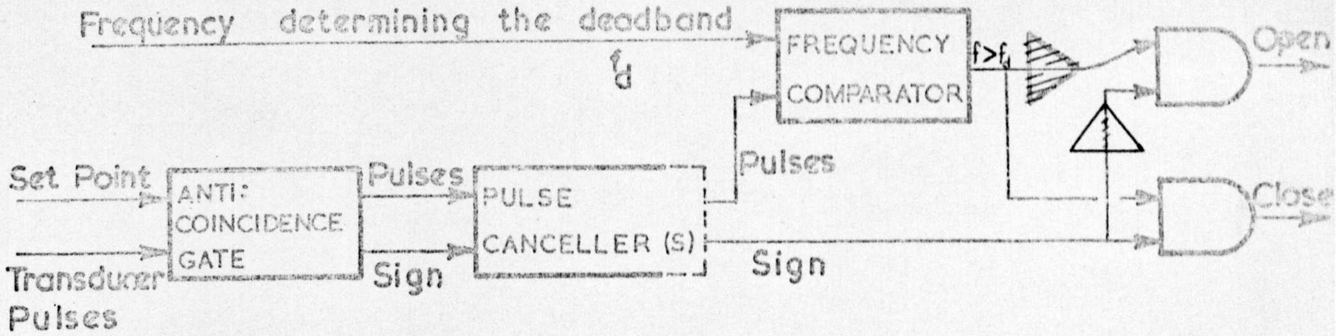
Fig 4B Comparison of normal and 'pulse-spilling' limiters incremental-spilling types of limiting by means of complex describing functions. Within a control loop, oscillations wherein the actuator moves back and forth over its entire range would usually be unacceptable. The slight extra attenuation, and more particularly the phase advance afforded by the spill mode of limiting, would give some reduction in the amplitude of such very large limit cycles. The property is of somewhat greater interest as a means of reducing the effect known as 'integral wind up' that causes excessive overshoots following large changes in controller set-points.

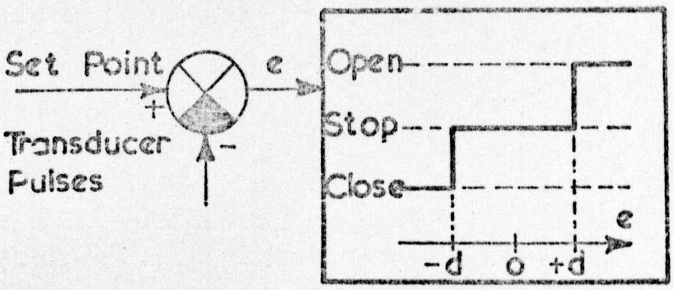
4.3.2 Two and three state controllers

The simplest form of controller, known as 'bang-bang', has a two-state output. In the pulse-frequency domain such controllers consist of an anti-coincidence gate, followed by one or more pulse-cancellers and a frequency comparator. The pulse cancellers are to remove the effects of jitter in the set-point and transducer pulse trains, which could otherwise result in a high frequency flutter at the controller output. Too many cancellers cause excessive hysteresis.

The controller with three output states (+, 0, -) is rather more complex. It is needed if the actuator design does not permit direct transition from one active state (e.g. valve opening) to another active state (e.g. valve closing) without passing through an intermediate passive state.

A design is shown below having two active states separated by a deadband. The error frequency ($\propto$ error modulus) is tested for size. If it exceeds a threshold (f_d) then the appropriate active output is generated as determined by error sign. Otherwise no output is generated.

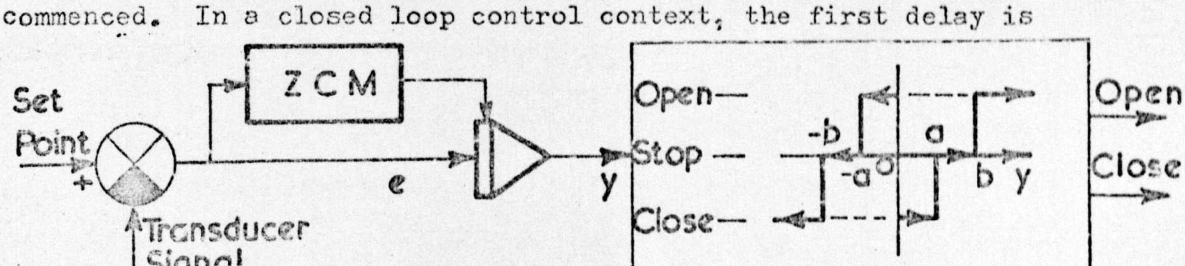




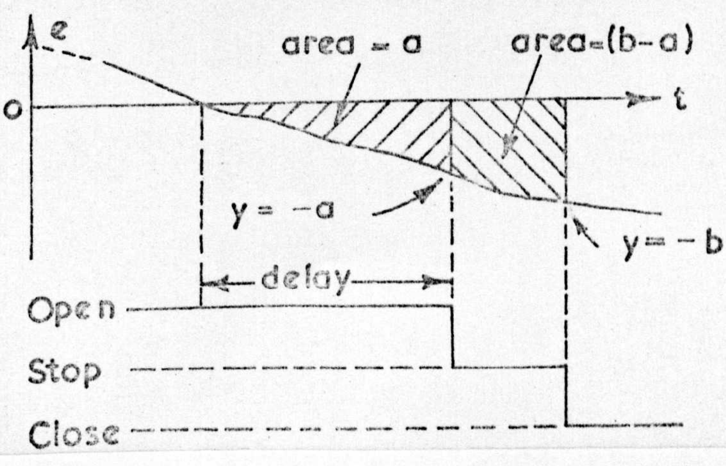
Ternary - Output Controller

Deadband = 2d

A simpler ternary output controller has a phase deadband, rather than a frequency deadband. That is, following a change in error sign, a certain number of error pulses must occur before the output changes. A block diagram indicating the working (but not the actual circuit) of a controller with phase hysteresis is shown below. The error (in fact a signed frequency) is shown going from positive to negative. Only after the time integral of the error (starting from the instant the error goes negative) has reached -a does the 'open' output command cease. Not until this integral reaches -b pulses is the 'close' output command commenced. In a closed loop control context, the first delay is

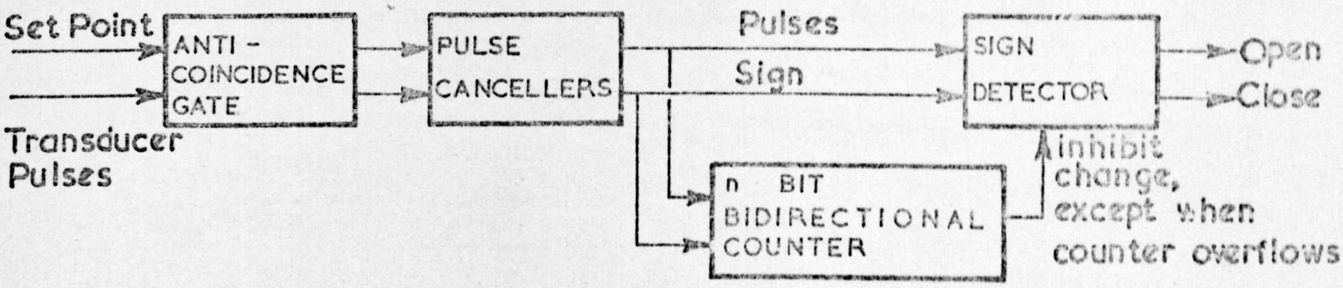


Note: the unit ZCM generates a pulse that resets the integrator whenever the error e crosses zero.



Ternary - Output Controller with Phase Hysteresis undesireable; during this delay, the control action is of the wrong sign. Ideally 'a' should be nearly zero, while 'b' should be chosen to compromise between good control (small 'b') and low actuator wear (large 'b'). Clearly the life of an actuator is greater if it spends a higher fraction of time in its inactive state.

The diagram below shows one circuit for a controller with phase hysteresis. The error pulses are passed through a number (c) of cancellers, to counteract input jitter, and then to a three state sign detector. The bidirectional counter is so connected as to effectively divide the pulse train out of the last canceller by a factor M ($M = 2^n$).

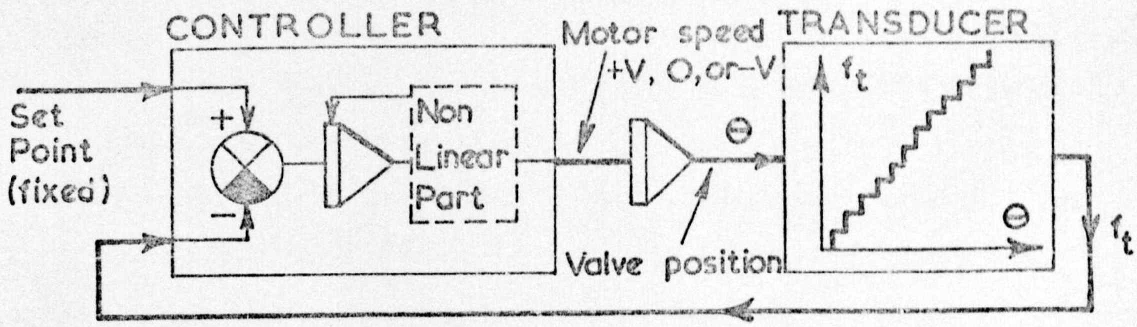


Ternary - Output Controller with Phase Hysteresis.

With this arrangement, 'a' has a value in the range c to c + M, depending on the count in the counter at the instant that the error changes sign. On average $a = c + \frac{1}{2}M$. 'b - a' = M. [By the use of more complex circuitry to reset the counter to 0 or to M - 1 as the error changes sign, it would be possible to reduce 'a' to the value c.] A three state controller, corresponding to the arrangement above, was constructed with 3 cancellers (c = 3) and a variable length counter, allowing M to equal 1, 2, 4, 8 or 16. The controller was coupled via relays to a diaphragm valve driven by a single-phase motor. The valve position was obtained, using a transducer already described, and fed back to the controller. The discrete nature of the wire-wound potentiometer in the transducer caused the transducer output frequency to be quantised into steps of about 16 Hz. The transducer range was 1kHz → 2kHz, so quantisation step = 1.6%. The valve motor was slow, moving the valve at a fixed speed of about 1%

of useful range per second. Motor over-run and transducer hysteresis were small, together constituting $< \frac{1}{2}\%$ of range.

The main features of the control loop are shown diagrammatically below:



The control loop was set up with a fixed set point, which happened to correspond fairly closely to one of the transducer step frequencies. The presence of two integrators and two non linearities in the control loop gave rise to oscillations of range ±1 transducer step frequency. The following measurements were obtained:

Set point frequency	1499.8 Hz
Mean transducer frequency (over 10^6 pulses ≡ 11 minutes)	1500.3 Hz
Mean error	- 0.5 Hz ≡ -.05% of range
Maximum error frequency	17 Hz
R.M.S. error frequency	13 Hz
Oscillation period	10 secs
Duration of 'motor stopped' state	1 sec ± 10%
Motion time/stopped time	4

The phase hysteresis was varied, and gave the following:

M = phase hysteresis	16	8	4
Maximum error frequency	17 Hz	17 Hz	17 Hz
Period of oscillation	10 sec	8 sec	5 sec
Stop time	1 sec	0.7 sec	0.3 sec (?)

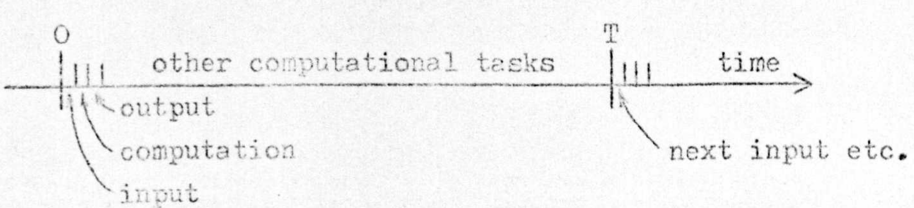
The minimum period would be about 4 seconds, being the time taken for the valve to traverse (twice) a transducer quantisation band. The most interesting feature of these results is the smallness of the mean error, at least as measured by the transducer. This is because for small errors the controller maintains a phase memory and is consequently acting as a phase-locked loop. The phase of the error is maintained within a band about 3M pulses wide, although the discontinuities in phase measurement (counter overflow) allow occasional additional phase shifts of M pulses.

The control situation just described requires almost continuous actuator movement, which is far from ideal, but arises from the paucity of actuators capable of responding to digital signals. The cycling action should be adjusted so that its period is markedly less than the dominant time constants of the controlled plant.

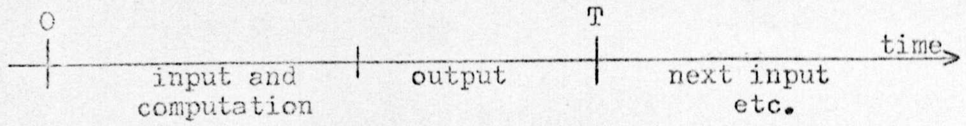
4.3.3 Sampling controllers

In the last subsection (4.3.2) controllers were described that have a pulse-train input, but an output that can take only two or three states. More complex controllers have a pulse-train output as well as a pulse-train input. These may be divided into continuous devices and cyclic devices. This subsection treats cyclic devices.

When a digital computer is used as a controller, the actions of (measurement) input, computation, and (actuator) output are performed sequentially. The sequence is repeated as often as the control situation is thought to require, for example every T seconds. As only a fraction of T seconds is needed for one cycle of the sequence, considerable time is available for other computational tasks.



A digital controller that uses counting techniques and p.f. signals is much slower than a computer using parallel arithmetic and binary coded signals. Consequently the time available for 'other computational tasks' is much less. The distinction between the three actions of input, computation and output is less sharp, because certain counting operations perform two of these actions simultaneously. A typical timing diagram for a counting controller is:



Counting techniques were described in Chapter 3. The ones most used in counting controllers are the following:

(a) Frequency to number conversion: A pulse train of rate f is counted into an initially empty counter. After time T_1 the counter contains $n_p = T_1\bar{f}$; Gain $= T_1$; Time $= T_1$.

(b) Rate of change of frequency to number conversion: A pulse train is counted <u>up</u> into an initially empty bidirectional counter for time T_2. It is then counted <u>down</u> for T_2.

$$n_d = T_2^2 \frac{df}{dt}; \quad \text{Gain} = T_2^2; \quad \text{Time} = 2T_2 .$$

(c) Addition of frequencies: Two pulse trains are fed into an anti-coincidence gate whose pulse output represents the sum.

(d) Addition of a frequency to a number: A counter is preset to the number, n_o, and then counted upwards for time T_1 at the given frequency. $n_s = n_o + T_1\bar{f}$; Time $= T_1$.

(e) Addition of two frequencies to give a number: A counter is preset to zero, then counted up for T_1 seconds at rate f_1, followed by a further T_1 seconds at f_2.

$$n_s = T_1(\bar{f}_1 + \bar{f}_2); \text{ Gain} = T_1; \text{ Time} = 2T_1 .$$

(f) Addition of two numbers: To add a number n_1 in register 1 to a number n_2 in register 2, pulses are counted out of register 1 into register 2, at any convenient clock rate f_r, until register 1 is empty

$$n_2(\text{final}) = n_2(\text{initial}) + n_1; \quad \text{Time} = n_1/f_r .$$

(g) Subtraction: All the methods of addition can, with slight modification, be used for subtraction.

(h) Multiplication: Any of the pulse trains mentioned in (a) to (g) above can be passed through a frequency scaling circuit.

(i) Integration: Any pulse frequency can be integrated by continuous counting. It is likely that the frequency will be signed, and the counter will be bidirectional.

Over the input and computation stages of its cycle, a counting controller builds up a number in a register. This number may be immediately output to an actuator, or it may be transferred into a short term memory (another register) where it is preserved until the following control cycle is complete. Immediate output is appropriate if the actuator is a stepping motor. Stored output is needed if the number is to be converted to a frequency, (or converted to analogue form).

While the stepping motor is a convenient actuator, the procedure of moving it, at the end of each control cycle, by a computed amount, leads to difficulties in controller design. A periodically incremented device is effectively an integrator. A 'proportional' controller, whose output is used as an increment, gives an integral action. A 'derivative' controller so used gives a proportional action. To obtain true derivative action requires a 'double derivative' controller, which is very difficult to construct.

The integrating properties of a stepping motor can be disguised by the use of differencing within the controller. If x is the value of the controller output, and y is the position of the motor (in steps from a datum), then it would be attractive to realise the relation $y = x$. Provided no pulses are lost, this relation is the same as $\nabla y = \nabla x$, i.e. the motor should be moved, each cycle, by the change in the controller output since the last cycle. Unfortunately to perform the subtraction $\nabla x = x_i - x_{i-1}$, in a way that does not destroy the number x_i, requires three registers and two transfer operations. Such complexity is out of proportion in a simple controller.

Any controller design must permit the control coefficients to be varied, and preferably to be varied independently. In counting controllers the coefficients are applied to scale pulse trains during input, internal transfers or during output. Where more than one mode of control is involved, it is not easy to prevent coefficients from inter-reacting, unless the modes are completely separated in time. For this reason controllers may be designed with alternating modes of action, each mode being a complete sequence of input, computation and output.

To illustrate the necessary complexity of a counting controller, the instruction sequence of a two term controller is listed below. The controller, which approximates 'P + I' actions, has the pulse transfer function:

$$\frac{Y(z)}{E_1(z)} = k_p + \frac{k_i}{1 - z^{-1}} = \frac{(k_p + k_i) - k_p z^{-1}}{1 - z^{-1}}$$

where $y(t)$ is the output stepping motor position, and $e_1(t)$ is a slightly delayed and filtered version of the measured error. The controller contains 4 bidirectional counters, of which A and B are long while C and D can be shorter. Three pulse trains are available in the controller: a slow one of rate f_s for driving the stepping motor, a medium one of rate f_c for timing, and a very fast one of rate f_f (= e.g. 10 MHz) for internal transfers. The controller set-point and sampling period are available as parallel coded numbers.

Examination of the instruction sequence shows up two features: that an elaborate sequence controller is required, and that counters capable of representing both positive and negative quantities are needed. These features are inter-related, because unidirectional counters may be employed at the expense of a more complex sequence control[313].

Instruction sequence of 2-term controller

n_s = set point

$N = \frac{1}{2} f_c T$ represents sampling period

$\Rightarrow$ indicates parallel loading of a counter

$+f \rightarrow$ indicates counting up at pulse rate f in the designated counter

^{1} indicates a value from the previous cycle

Nº	Action	Stop when	Time	Counter contents A	B	C	D*	Motor position
0	Initial state			0	0	x_p^1	r^1	y^1
1	$n_s \Rightarrow A$; $N \Rightarrow B$	—	0	n_s	N	"	"	"
2	$-f \rightarrow A$; $-f_c \rightarrow B$	B = 0	$\frac{1}{2}T$	$n_s - \frac{1}{2}fT = e$	0	"	"	"
3	$-f_f \rightarrow A$; $+k_p f_f \rightarrow B$; $+k_i f_f \rightarrow D$	A = 0	$\doteq \frac{1}{2}T$	0	$k_p e = x_p$	"	$r^1 + k_i e = r^1 + x_i$	"
4	$-f_f \rightarrow C$; $-f_f \rightarrow D$	C = 0	$\doteq \frac{1}{2}T$	"	"	0	$r^1 + x_i - x_p^1$	"
5	$-f_f \rightarrow B$; $+f_f \rightarrow D$; $+f_f \rightarrow C$	B = 0	$\doteq \frac{1}{2}T$	"	0	x_p	$r^1 + x_i + (x_p - x_p^1) = X_i$	"
6	$N \Rightarrow B$	-	$\doteq \frac{1}{2}T$	"	N	"	"	"
7	$-f_c \rightarrow B$; $-f_s \rightarrow D$; $+f_s \rightarrow$ motor	D = 0 go to 8						
		B = 0	$\doteq T$	0	0	x_p	r	$y^1 + X_i - r$
8	$-f_c \rightarrow B$	go to 1						

* Note that the residue left in D from the previous operation (r^1) or this operation (r) will usually be zero. Non-zero residues result if the motor is still moving at the end of the cycle.

4.3.4 The Serck Controller

In 1966 the Serck Company developed a two term digital controller. A contract was placed with the University of Warwick for the analysis of its performance. The first year of research leading up to this thesis was devoted to that analysis. A long report was made to the Company, and with their permission a conference paper was prepared from it (IFAC Budapest 1968). The conference paper was considerably compressed for publication in Automatica in 1969. Both* the conference preprint and the paper are included at the end of the thesis: the latter contains the core material in the more readable form. In this section only the salient features of this rather special controller will be summarised.

The controller performs proportional and integral action by counting techniques. It is consequently a member of the class of sampling controllers considered in the last sub-section (4.3.3). Frequency differentiation is performed by the count-up count-down method. The result of the differentiation is passed to a stepping motor through a frequency scaling circuit, resulting in overall proportional control. Alternating with the proportional control is integral control, involving counting (to obtain the error) followed by transfer to the stepping motor.

The distinctive features of this controller may be best seen by comparing it with a digital computer programmed to perform P + I control. The Serck controller:

(a) Has a cycle time varying with the measured variable. The transducer employed gives a frequency inversely proportional to temperature, and each stage of the control cycle commences with counting a fixed number of transducer pulses.

(b) Has a cycle time varying with both error and rate of change of measured variable. The size of the increments to the stepping motor output depend on these two quantities; so does the time to apply them. This 'runout' time prolongs the cycle (or 'sampling') time.

* Correction - only the paper has been included.

(c) Applies proportional and integral control actions alternately, rather than simultaneously.

(d) Generates a proportional action solely upon knowledge of the rate of change of the measured variable. The name 'pseudo-proportional' seems appropriate to this action. It differs from conventional proportional action, which is calculated upon knowledge of the measurement error (= set point minus measurement).

(e) Exhibits filtering and delay (that belong to counting methods of frequency-to-number conversion) when compared with rapid analogue-to-digital conversion.

(f) Is particularly prone to quantisation errors on account of the difference-and-increment method of obtaining proportional action. This aspect is dealt with in a later section.

The controller was built at a time when logic integrated circuits were still expensive. There was thus good reason to try to minimise the number and length of registers. The distinctive features of the design arose primarily as a result of such minimisation, rather than in pursuit of a novel control algorithm.

In the context of chemical process control, for which the controller was designed, many of the distinctive features caused negligible variation from conventional analogue P + I control. In experimental work, this latter was taken as a yardstick. Chemical processes have long time constants, and there was little difficulty in arranging the controller to perform many cycles during one time constant. The ratio V of dominant time constant to cycle time is a useful measure. It was found by calculation and experiment that for values of V above about 6 the effects of (c) and (e) above were negligible.

Distinctive feature (d), the 'pseudo-proportional' action, does not affect loop stability or the power to reject disturbances, when compared with true proportional action. It does affect the response to a change in set point. This response is determined solely by the integral action, and is rather sluggish. It is an interesting question of control practice, whether step changes in set-points should be a common phenomenon. Certainly good step response is often only obtainable at the expense of sub-optimum performance in rejecting disturbances.

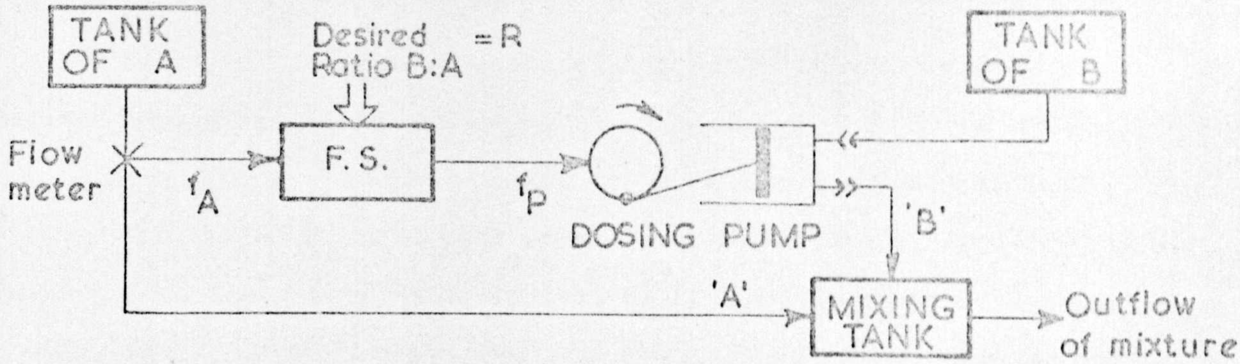
Features (a) and (b), variation of cycle time as a complex function of measured variable, were a cause for concern. Various mechanisms for instability were envisaged yet calculations failed to yield any usable expressions. Extensive hybrid computer simulation was undertaken, in which the variations were exaggerated, but no tendencies to instability could be found except at very slow sampling rates, e.g. $V < 3$. The principal effect of cycle time variation is to induce a corresponding variation in the integral action gain. A secondary effect is to increase computational delays.

Commercial convenience favours a controller with the minimum number of adjustable parameters. The sampling rate of a digital controller is an extra parameter, having no counterpart in analogue controllers. Considerable attention was given to determining the performance of a controller with a fixed sampling rate applied to processes with widely differing time constants, i.e. the ratio V was varied over a wide range. At low V the various skew sampling and delay features cause a decline in stability margins. At high V, it is quantisation effects that cause poor control.

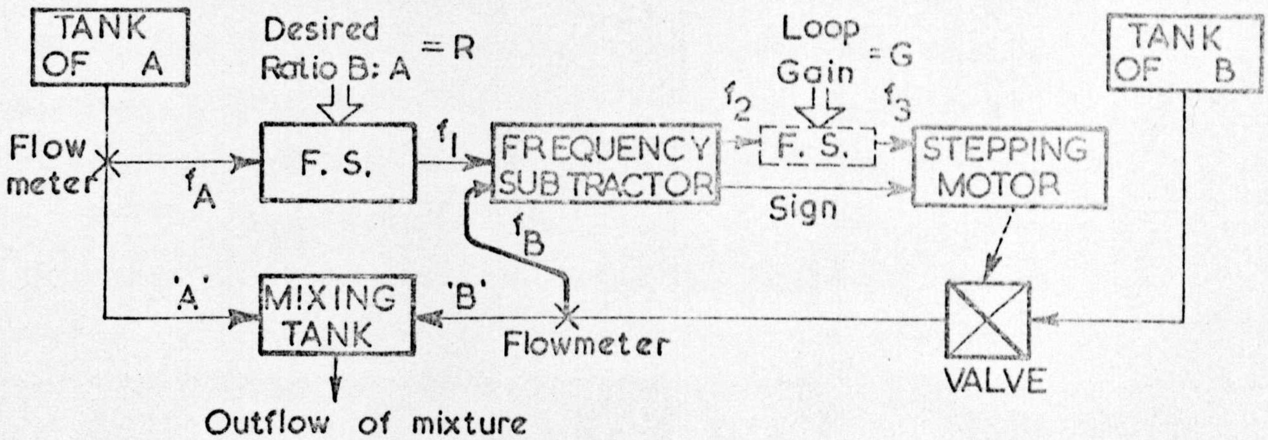
Attempts, while evaluating the Serck Controller, to write down exact equations of behaviour, proved unprofitable. Such notations as the delayed z-transform gave elaborate expressions that cound not be generalised, whereas particular cases could be more easily examined by simulation. It is clear that the calculation of the settings of a two-term controller in an industrial application should not require knowledge of complex notations.

4.3.5 Continuous controllers

A continuous pulse-frequency controller is simpler to understand than the sampling ones of the last section, for it requires no sequencing logic. A continuous controller has a smoother action than a sampling one, but its construction usually requires more components.



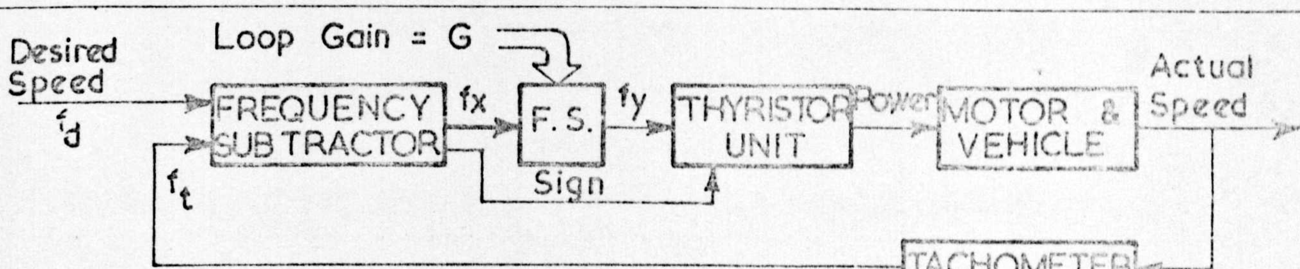
(a) 'Open loop' control of fluid ratio



(b) 'Closed loop' control of fluid ratio

Notes: F.S. is a frequency scaling circuit

$f_P = R.f_A$; $f_1 = R.f_A$; $f_2 = |f_1 - f_B|$; $f_3 = G f_2$



(c) 'Closed loop' control of motor speed

$f_x = |f_d - f_t|$; $f_y = G\ f_x$: Power $\propto$ (Sign). f_y

Fig. 4C Simple continuous P.F. controllers

The principal construction blocks are coincidence gates with cancellers (for addition and subtraction), bidirectional counters and the frequency scalers. Auxiliary blocks are frequency comparators, reference oscillators, and sign logic. Where the control set-point represents the desired value of a variable it is most easily expressed as a frequency. Alternatively, where it represents the desired value of a ratio of two variables, numerical form is most convenient. Number-to-frequency conversion is generally simpler than frequency-to-number conversion.

Three simple controllers are shown in Fig. 4C. They all operate continuously on PFM signals from a transducer and generate an output pulse train. Controller (b) has an integral action because of its stepping motor actuator. Its loop gain can safely be set to its maximum value (K = 1) unless there is a long delay between moving the motor and observing changes in flow rate. A machine tool version of this controller is described by Ladanyi.[331] Controller (c) has a proportional action. Here the useable gain depends upon the motor characteristics, but it is likely to be large. All three controllers have only a single action, so there are no difficulties with sign. However the inability of a standard frequency scaler to increase a frequency limits the controller gains to 1, which may be well below the optimum.

Fig. 4C (a) shows an open loop controller. Other open loop continuous controllers (handling PFM signals) are applicable to machine tools. Numerically controlled machine tools are increasingly being fitted with stepping motors[419,420,421], as the latter become available in higher powers. The tools are moved according to coded instructions stored on magnetic or paper tape. For point to point operations, counting techniques are used to transfer a coded instruction (e.g. advance tool by 15.60 mm.) into motor movements. For contouring operations a more sophisticated motor controller is required that co-ordinates movements on different axes, maintains a particular speed of movement, performs co-ordinate conversions or similar operations.

The simplest contouring motion is the following of a straight line. If the line is at angle θ to the x axis of a Cartesian co-ordinate system, then the motor velocities on the y and x axes should be maintained in the ratio tan Θ:1. This could be achieved by feeding a pulse train of rate f to the x axis motor while feeding the same train to the y axis motor via a frequency scaler set to tan Θ. A better arrangement, allowing direct control of actual speed, is the application of the pulse train through separate scalers to the x and y axis motors. If the scale factors are cos Θ and sin Θ respectively they can be realised for all Θ, as $|\cos\Theta|$ and $|\sin\Theta|$ never exceed unity. Moreover the resultant speed is proportional to f, which may be adjusted according to the task at hand, viz. engraving, flame cutting, milling etc. The coefficients cos Θ and sin Θ applied to the two scalers will be in number form. If necessary they could be formed from the angle Θ by counting techniques (Section 3.9.3), even employing the scalers themselves for this operation provided their outputs are disconnected from the motors.

Other contours than straight lines can be generated, using polynomial and exponential function generators. Whatever the contour, there will be rounding effects, which are predictable but complex. For straight lines these will not exceed 1 motor step, but for curved lines they may be greater. A machine tool obeying a large number of essentially incremental instructions, each subject to a small rounding error on implementation, may accumulate a sizeable error. This may be avoided by pre-calculation of rounding error, or by some element of feedback.

Multi-term continuous controllers require the merging or cancelling of a number of pulse trains. As described in Section 3.2, there is a choice of synchronous working or asynchronous working. The latter requires liberal use of the relatively complex coincidence gate. The former creates severe design problems, as all pulse trains must be synchronised with particular phases of the clock in a way that prevents conflicts. For mass-produced devices, synchronous working is usually cheapest. Where only a few controllers are to be made, the saving in components is more than offset by the increased difficulty of design; here, asynchronous working is recommended.

The handling of sign was treated in section 3.6. Following addition or subtraction, logic operations must be performed to determine the sign of the output and which of two output pulse trains should be used. The appropriate logic tables are shown in Fig. 3Q, and may be thought of as a 6 input, 2 output structure, viz.

Inputs:	sign of first input	Outputs:	sign
	sign of second input		pulse train
	sign out of canceller		
	operation (+ or -)		
	pulse train from coincidence gate		
	pulse train from canceller		

As is often the case, there is a choice between minimum hardware and minimum design complexity. An integrated circuit 'read only memory' could be attached to every coincidence gate and canceller, converting them into universals signed adder-subtractors. Such a circuit might not be cheap, but it would be very simple to use. Alternatively every addition or subtraction can be examined separately, so that by using all information concerning possible signs and amplitudes of the inputs, the minimum logic is specified.

A practical problem with all digital controllers is the setting of the (numeric) control coefficients. Elaborate switches are expensive and bulky, yet there is no digital equivalent of the simple analogue potentiometer. Where coefficients have a limited range, standard binary or decimal coded switches are employed. Where the coefficients have a wide range (e.g. 3 decades), fully coded switches are wasteful, because they do not give the sort of logarithmic scale of increments required. The simplest logarithmic scale that can be implemented is that comprising the powers of 2. A single armature n-position switch can be connected to a binary frequency scaler (e.g. BRM) to give a scale of coefficients: 1, 2, 4, ..., 2^{n-1}, i.e. a logarithmic increase of 100%. With binary coded decimal scalers, the simplest scale is: 1, 2, 4, 10, 20, 40, 100, etc., again achievable using a single armature (single 'pole') switch. Where logarithmic increases of less than 100% are needed, multi-armature switches are needed, capable of applying logic '1' signals to more than one stage of a scaler. The table below indicates some arrangements for binary scalers:

Steps per octave	1	2		3		7
Mean % increment	100%	41%		26%		10½%
No. of switch poles	1	2	3	2	3	3
Max. % increment	100%	50%	45%	33%	30%	14%
Notes	1-2	2-3-4	8-11-16	4-5-6-8	8-10-13-16	8-9-10-11-12-13-14-16

Unlike coefficients, set-points have to be specified very precisely, and again digital devices present special problems. For manual settings by semi-skilled operators, decimal rather than binary coding is almost mandatory. However even using (binary-coded) decimal scalers, difficulties result if the set-point range is constrained to be a power of 10. A temperature controller for the interval 0-200°C is not convenient if provided with a set-point scale of 0-1000. What with analogue devices requires only a changed label or printed scale, with digital devices requires a change in internal circuit. Fortunately one decade of a decimal rate multiplier (DRM) can be rapidly modified to act as scale-of-two or scale-of-5 multiplier. So a 3 decade DRM can be used to give scales of 0-999, 0-499, and 0-199. Other variations, such as 0-1.99 or 200-399 can be handled by relabelling, as with an analogue device.

4.3.6 Heirarchical control systems

Most control systems can be considered as having a heirarchy of functions. At the lowest level are very simple functions that need to be performed continuously, often in a number of places simultaneously. At the highest levels, the functions are complex and system-wide; they are usually performed discontinuously and are not urgent. A batch process chemical plant has a well ordered control heirarchy as follows:

Level	Function	Typical Frequency	Agent
1	Actuator movements	Continuous	A
2	Single loop control to a set point	"	A
3	Generation of set points (e.g. start up)	Every 20 mins.	A or H, O
4	Determination of controller parameters	Every 2 hours	A or H, O
5	Scheduling of production	Every 3 weeks	H, M
6	Configuration of control system	Every 6 months	H, D
7	Identification of process characteristics	Every 1 year	H, R
8	Choice of manufacturing method	Every 5 years	H, M

A = Automatic device; H= Human being

O = Operator; D = Designer; M = Manager; R = Researcher

There is some overlap and interaction of these functions, but in general each function communicates with one higher level function and several lower level ones. Automation is only easily applicable to the very lowest levels of activities; human intelligence is required for the higher levels. However the transition point from the former to the latter has been slowly rising.

The cost of communication between one level and another is high and sometimes constitutes the major cost of a control system. The most complex of these interfaces is the one corresponding to the transition from automation to human operation. The trend to centralised control is motivated primarily by the need to reduce the cost of transferring information to and from a plant operator. Centralised control has other advantages, such as ease of instrument maintenance, and greater instrument reliability.

Computer process control is an extension of centralised control whereby a single computational processor is shared between a large number of tasks at more than one level. Some of the inter-level communication is performed by means of internal transfers within the computer. Computer control is usually implemented using versatile general purpose machines, which are mass-produced and hence relatively cheap. However these machines have often proved disappointing for the following reasons:

- Interleaving tasks of different levels of complexity and urgency requires a complex programme. Operators do not understand such programmes and cannot modify them.

- A computer control system has little redundancy, so that all parts require high reliability. Central processor failure results in simultaneous loss of control of all variables - a condition so hazardous that special 'stand-by' controllers are usually necessary.

- The costs of the computer-process and computer-man interfaces are high, as are those of programming a particular set of tasks.

- Computer control is particularly difficult to commission. Two machines, rather than one, are really needed. It is not always possible to build up a control system one task at a time.

- Manufacturers' emphasis on hardware has sometimes obscured the importance of software preparation, causing expensive delays in commissioning.

In many control applications, the versatility of a general purpose computer is not needed, and only contributes to the difficulty of understanding it. The association of a particular circuit with a particular and unchanging function is the negation of this design versatility. However, it is attractive for control systems, provided the circuits can be made cheaply enough. The term 'firmware' (= hardware + software) has been used for this sort of association, which reflects a trend in manufacturers' thinking[418]. The high cost of components has resulted in a number of compromise designs, usually known as multi-loop controllers, which are particularly applicable to the control of medium size processes (e.g. 10 loops) or automatic testing stations.

Two British designs[321,422] of multiloop controller appeared on the market in 1970. Both employ a shared central processor having a small repertoire of special instructions, linked to a number of autonomous input and output circuits. Because the central processor is digital, the peripheral units are also digital. However the latter use counting techniques while the former uses faster parallel coded techniques. Transfers between the centre and the units is by counting or shift operations over single wires. The output units are so arranged that output signals continue even during a central processor failure. These multiloop controllers perform functions of levels 1 and 2 of the table above, are capable of some extension into level 3, and also organise the interface with the operator to his convenience.

A full development of the techniques described in this thesis would permit a considerable extension of digital control, using a heirarchy of devices corresponding to the heirarchy of functions. Pulse-frequency signals offer a relatively cheap form of communication between devices; pulse-counting techniques can be used for all but the highest level functions; the computational and communication tasks are thereby properly and simply integrated. The association of a distinct and autonomous device with each task simplifies commissioning, permits piecemeal elaboration, and gives operational security. On the other hand, the use of a standard digital signal simplifies the superposition of higher level control or monitoring devices. These latter, with their more complex functions could well be general purpose computers.

An autonomous valve-positioning controller was described in section 4.3.2. A two-term (P + I) controller to supervise the valve positioner was developed and is described in the following section. These two devices carry out functions at levels 1 and 2 respectively, and have been operated under the direction of a g.p. computer. At the present state of technology special devices for level 3 functions would probably cost more than a computer.

4.4 A two-term process controller

4.4.1 General description

In order to demonstrate the principles of a continuous pulse-frequency device, and to test many of the circuits described in Chapter 3, a two-term process controller was built. Its general specification was taken from current practice with analogue controllers, that is:

(a) The two inputs and one output are to a common scale

(b) Proportional and integral actions are provided, each one adjustable. The transfer function in Laplace transform notation is

$$X(s) = k_p\left(1 + \frac{k_i}{s}\right). E(s)$$

where $x(t)$ is the output, $e(t)$ the difference between the inputs.

(c) The set point input can be generated locally or received from a remote device.

(d) The set point and measurement inputs, the error and the output can be observed by an operator.

(e) Integral 'wind-up' is prevented; i.e. the integral term does not continue to increase after the output reaches its limit.

In addition to these standard features the following two were incorporated:

(f) Resolution and steady state error are better than ½%

(g) The set point input is continuously tracked, and preserved in a digital memory. This permits control to a set point that is only intermittently specified.

A block diagram of the controller is shown overleaf, while Photo B shows its front panel.

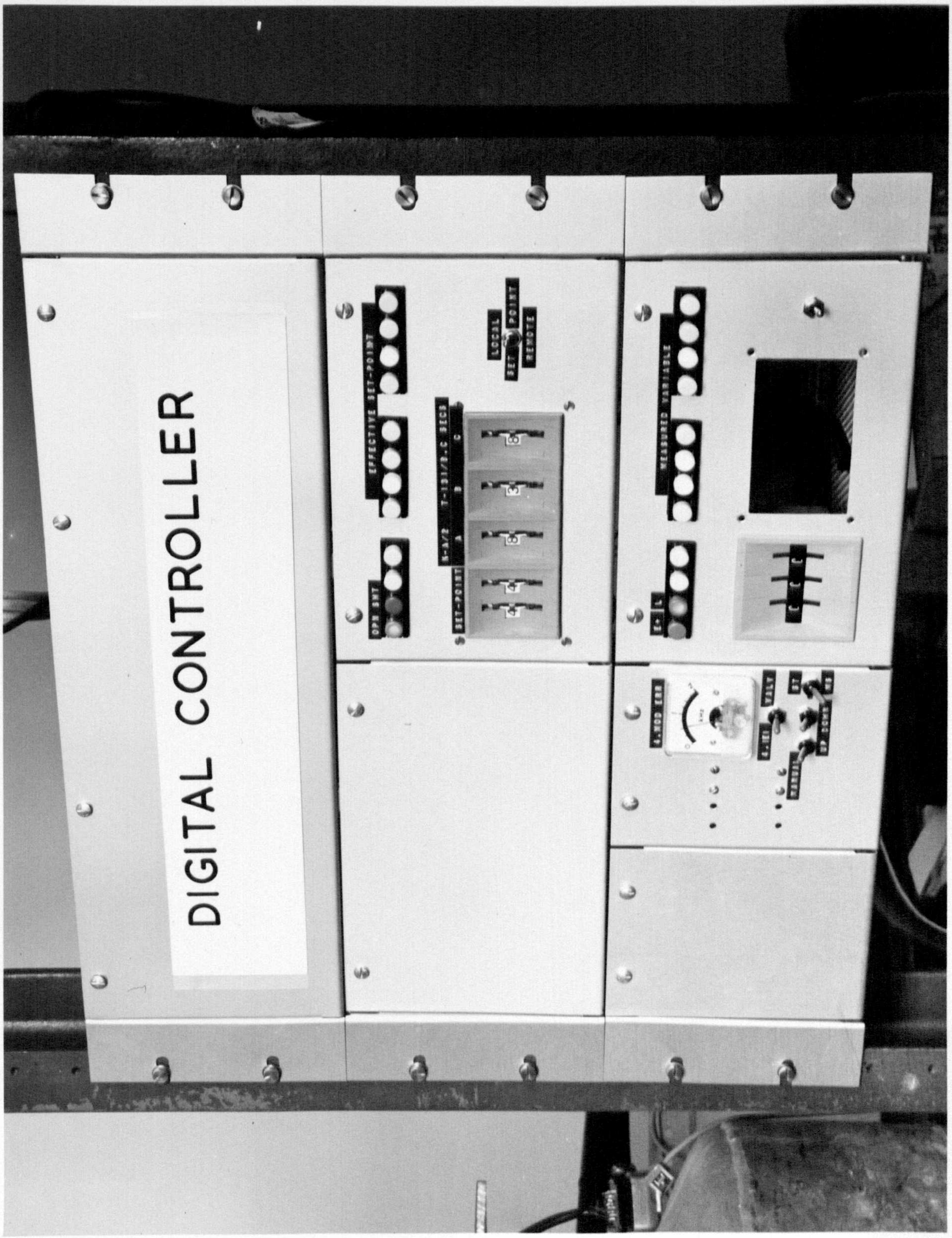


Photo B: Front panel of two-term continuous controller.

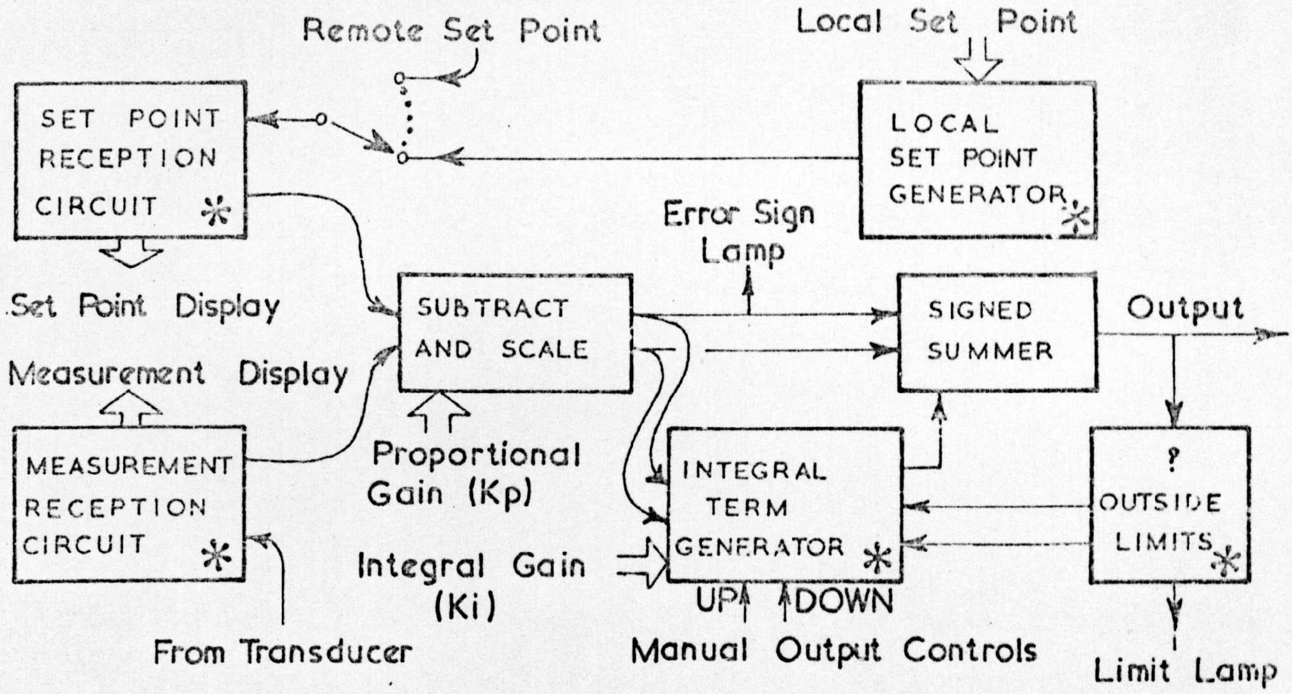


FIG. 4 D BLOCK DIAGRAM OF TWO-TERM PROCESS CONTROLLER

To simplify construction, binary rather than decimal coding has been used throughout the controller. Thus the two digital displays are 8 bit binary, while the local set point is determined by two 16 position thumbwheel switches, i.e. 8 binary bits. As $2^{-8} = 0.004$, the resolution criterion of 0.5% is just satisfied. Although a number of devices are marketed with inputs or displays in binary code, decimal coding is strongly preferable. A major attraction of a digital display is the ease with which it can be read, and the consequent reduction in misreadings. A requirement to convert readings from binary to decimal code quite outweighs this attraction.

4.4.2 Design details

Schematic and connection diagrams for the various parts of the controller are given in Appendix A, where circuit construction is also described. The purpose of this section is to comment on the block diagram of the previous page. That diagram shows the general flow of single wire signals by lines and of parallel coded digital signals by broad arrows. Also, all the blocks requiring a reference (clock) pulse train are marked with an asterisk.

The reception circuit for the measurement input subtracts the standard bias (of 1kHz) from the transducer frequency, multiplies the residue by 16 using a tracking loop, and synchronises the pulses with one phase of the main clock. The digital display of measurement is obtained from the register in the tracking loop. A 1-bit hysteresis has been introduced (see Chap. 3.7.3) to make the displayed number jitter-free. A signed frequency associated with the rate of change of the measurement is also available from the tracking loop, but was not used in this (P + I) controller.

The reception circuit for the set-point input is very similar to that just described. A different clock phase is used. The anti-jitter feature was omitted, but an auxiliary circuit is provided to test the validity ($>$ 1 kHz frequency) of the set-point input. Whenever this input becomes invalid (e.g. disappears), the tracking loop is inhibited so as to preserve its current state and act as a set point store. This feature is an important one if the controller is to be reliably operated under computer supervision.

The local set point generator is a binary scaling circuit. As the set point reception circuit contains a display, the local generator could have been a coarsely calibrated variable oscillator. In that case an operator would vary the oscillation frequency until the correct set point was displayed, then disconnect the oscillator.

The amplified inputs are carried by non-coincident pulse trains, which are subtracted (cancelled) and scaled by the proportional setting, K_p. The error signal so obtained is a signed frequency in the range -15kHz to +15kHz, as the scaler multiplies by a 4 bit binary fraction between 0 and 0.1111. To prevent unneccessary saturation of later stages, the error frequency modulus is now limited to 2kHz.

The integral term generator consists essentially of a bidirectional counter controlling a scaler to give a frequency in the range 2kHz to 4kHz. The bidirectional counter is preceded by other dividers and scalers whose configuration and setting determine the integral gain K_i (units of $secs^{-1}$). Direct control of the output is also available via the integral term generator. An operator command to increase the controller output causes the counter to be counted up (at 62 Hz), overriding its usual mode of operation. A further override is used to prevent the controller's output going outside the limits of 1kHz and 2kHz. When the output frequency reaches about 2.03 kHz, the counter in the integral term generator is decremented until the output $<$ 2kHz. Excursions below 1kHz are similarly treated.

The output stage of the controller combines the integral term (a frequency in the range 2 to 4kHz) with the proportional term (range -2 to +2 kHz) and then divides pulse frequency by 2 in order to reduce jitter. The output range could then be 0 to 3 kHz but for the action of the action of the limit detectors in bringing it within 1 to 2 kHz, as required.

The maximum proportional gain is $16 \times 0.1111 \times \frac{1}{2} = 7\frac{1}{2}$. The maximum integral gain is about 1.1 sec^{-1} and is variable down to about 0.01 sec^{-1}. A wider range would be useful for some slow industrial processes.

4.4.3 Performance

After a number of modifications and adjustments, the controller worked correctly. Intermittent faults due to unusually large noise spikes, or to imperfect edge connector contacts, continue to occur, but at intervals of several hours.

The set point reception circuit was driven from an external oscillator which was suddenly disconnected. The set point display fell on average about $1\frac{1}{2}$ bits $\equiv 0.6\%$ before freezing its value. This was thought to be satisfactory, but could be improved if necessary be extending the register in the tracking loop.

The measurement display was flicker free, as intended, and in contrast to the set point display.

For a fuller examination of controller performance, the controller was used to regulate pressure in a vessel. The pressure vessel which contained air in its upper part, was supplied with water from a pump, via a motorised valve ('the actuator'). Water flowed out of the vessel through a constricted pipe (containing a flowmeter) back to the reserve tank from which the pump was drawing. The actuator was operated by signals from the three state controller already described (Section 4.3.3), which drew its set point from the output of the P + I controller now under discussion. A general view of the apparatus is given in Photo A, with the pressure vessel on the left and both controllers in the rack on the right.

From a control point of view, the apparatus is very non-linear, and also subject to drift in characteristics. The actuator is velocity limited, as well as position limited. High gain control loops consequently show instability at large errors, but stability at small errors.

The transfer function relating valve position to tank pressure was measured as approximately $2.5/(1 + 300s)$ at a particular part of the pressure range (around 3 atmospheres pressure). Using a complex describing function to account for velocity limiting in the actuator, the control loop was calculated to be unconditionally stable provided the integral gain satisfied the inequality

$$K_i \times 300 < 1 + \frac{1.23}{2.5\,K_p}$$

or approximately

$$K_i < 1/300 \quad secs^{-1}$$

Using higher gains would only lead to oscillations if sufficiently large transients had excited the actuator into its velocity limited state. The lower graph of Fig. 4E shows a Nyquist diagram of the control loop under various control settings, while the upper graphs show the responses to small steps. The settings used were:

(i)	Controller gain, K_p = 4	Reset time, $1/K_i$ =	65 secs
(ii)	" 4	"	32 secs
(iii)	" $7\frac{1}{2}$	"	32 secs
(iv)	" $7\frac{1}{2}$	"	16 secs
(v)	" $7\frac{1}{2}$	No integral action, $K_i = 0$	

Examination of the step responses suggests curve (iv) is close to instability, having overshot by more than 100%. The Nyquist characteristics for the linear part of the loop under setting (iv) crosses the actuator's describing function at $\omega = .055$. The actuator velocity traverses full range in 70 seconds. So instability would be expected if the amplitude of actuator movement exceeds a sinusoid of peak value: X x velocity limit/ω = 2.1/(70 x .055) = 0.55, or just over half its range. With the small step applied, the actuator did not move more than about 20% of its range, and the response is just stable.

A more direct test of whether a controller is acting in the expected way is the comparison of calculated and measured step responses. Unfortunately even the small (6%) step responses shown in Fig. 4E indicate marked non-linearities. The positive and negative steps over the same range are quite different from each other at the higher gains: curves (iii) and (iv). The asterisk on both graphs indicates the position of the calculated overshoot peak of curve (ii). The calculation did not allow for the actuator's velocity limiting. Calculated and measured responses are close enough to confirm that the controller has approximately its designed performance.

Even with the integral gain set to its minimum value of about

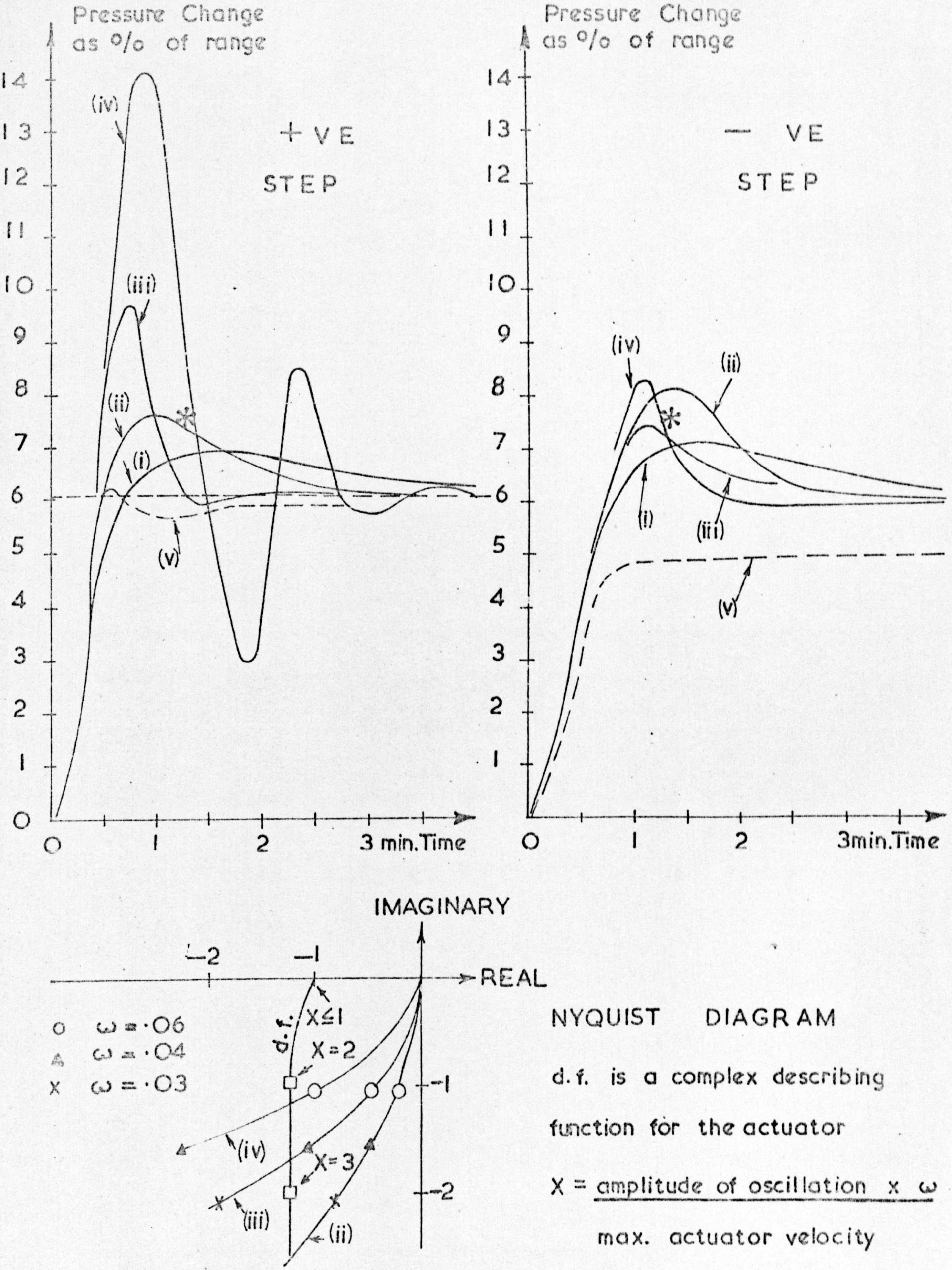


Fig. 4 E Step responses and Nyquist diagram for various Controller settings.

$0.01\ sec^{-1}$, the steady state error was less than 0.2% over several experiments. So the controller does not introduce any significant offsets.

Finally, the controller was operated as a slave to a computer, the latter generating set points and logging such variables as vessel pressure, flowrate out of the vessel, actuator position and demanded actuator position. For this test the computer analogue outputs were used for generating the set point signal (pulse train), while interrupts were employed for the inputs. Although this worked satisfactorily, it was an inefficient way of using the computer's central processor. In a permanent installation other forms of interfacing, as discussed earlier, would be preferable.

The controller has been shown to be a reasonable replacement for a two-term analogue device. It has high accuracy and stability (e.g. clock drift over 3 years $< 0.1\%$), an efficient method of avoiding 'integral wind-up', and with a little further development could have decimal numerical displays. It is suitable for cascade connection and is adapted for use in multi-level automation.

On the debit side, it is very bulky and is expensive of components. A fully developed design would contain about £150 of components at 1971 prices, mostly in the form of from 50 to 100 integrated circuit packages. Both price and package count are expected to continue to fall, but it is unlikely that a digital controller will ever be cheaper than an analogue one.

4.5 Accuracy and stability

4.5.1 Accuracy

The accuracy of a control scheme that uses analogue devices depends upon the quality of conversion standards, upon the linearity and stability of computational and storage elements, and upon the degree of interference in communication channels. Digital devices are subject to a different error mechanism, that of rounding [340,423]. P.f.m. control systems may exhibit both analogue and digital types of error, but these are generally small.

Long term computational errors are negligible with pulse frequency techniques, while pulse frequency modulation has good transmission properties. It is therefore the conversion from one signal form to another that most commonly introduces drift into p.f.m. control loops. Digital to frequency conversion requires only one standard. As this is usually the frequency of a fixed oscillator it is obtainable with excellent stability. Analogue to frequency conversion, by contrast, requires two standards : one each for the analogue quantity and for the frequency. For example, in a particular voltage to frequency converter, the former standard is a zener stabilised voltage threshold, while the latter is the time constant of a resistor capacitor combination. Such standards are difficult to stabilise against small drifts (such as 0.1%).

Cyclic controllers, such as the Serck controller described in section 4.3.4, involve both frequency to number conversion and rounding during computation. Errors occuring during the former were discussed in section 3.1.6. and shown to have zero mean value and a variance of $I^2/6$. (I is the resolution of the converted quantity in its digital form). Errors caused by rounding depend on the type of actuator and indirectly on the cycle time T. Cyclic controllers generally feed stepping actuators; a number of pulses is passed from the controller to the actuator at the end of each cycle. The computed actuator increment will generally not be exactly a whole number of pulses, but a number $X+x$, where X is integer, and x is fractional. Two strategies are possible - either the fractional part x is stored and added to the output for the following cycle - or it is disregarded. The former strategy complicates controller design. The consequences of the latter strategy will now be considered.

An exact analysis of a control loop with rounding is laborious and requires a full knowledge of the characteristics of the controlled process. Two approximate approaches complement each other in giving a measure to the effects of rounding. One is to calculate 'deadband' size; the other is to obtain statistics of 'quantisation noise'.

Deadband is best illustrated by the particular case of a cyclic controller giving an output increment proportional to the deviation of a measured variable from a set point. This corresponds to integral action, and for a given integral gain the controller gain K is proportional to the cycle time, T. If the deviation of the measured variable is less than $1/K$, then the output increment will be less than 1 actuator step and will be lost due to rounding. The deviation range $\pm 1/K$ is called a deadband, and it increases with the sampling rate $1/T$.

Following a change in set point, or some other disturbance to the process, the input to a controller will not lie in its deadband. Output increments of more than 1 step will be generated, which will acquire due to rounding an error of q (where $0 < q < 1$, or $-\frac{1}{2} < q < \frac{1}{2}$, according to the rounding rule used). Although the rounding action is deterministic, the rounding errors have statistical properties that are largely independent of those of the output increments from which they arise. Under some circumstances [333-335] the error sequence may be regarded as one of random quantities having the properties:

$$-\tfrac{1}{2} < q < \tfrac{1}{2} \; ; \quad p_q(z) = 1 \, , \text{ provided } -\tfrac{1}{2} < z < \tfrac{1}{2} \; ;$$

$$\langle q \rangle = 0 \; ; \quad \langle q^2 \rangle = 1/12 \; ; \quad \langle q_j . q_k \rangle = 0 \, , \text{ if } j \neq k \; ;$$

where q represents any rounding error, q_j that at the end of the j^{th} cycle.

Under these circumstances, the actuator movements are the sum of the proper computed increments, and of a train of random error steps of zero mean amplitude. In the absence of feedback, these latter would cause the actuator to perform a 'random walk' or Markov process. After n steps the statistics of the cumulative error would be:

$$-n/2 < Q_n < n/2 \; ; \quad \langle Q_n \rangle = 0 \; ; \; \langle Q_n^2 \rangle = n/12 \; .$$

The presence of feedback will eventually compensate this actuator error. Let the response of the measured variable to a single actuator step be denoted $h(t)$. If the control is of Class I or higher, $h(t)$ will have a finite energy and $h(t) \rightarrow 0$ as $t \rightarrow \infty$. The error in the measured variable attributable to the controller's output rounding error (q_i) occuring i cycles earlier will be $q_i . h(iT)$. The total error in the measured variable due to rounding will therefore be:

$$e = \sum_{i=0}^{\infty} q_i \; . \; h(iT) \; .$$

As the q_i are uncorrelated, the mean and variance of e are:

$$\langle e \rangle = \langle q \rangle \sum_{i=0}^{\infty} h(iT) \; = \; 0 \; .$$

$$\langle e^2 \rangle = \langle q^2 \rangle \sum_{i=0}^{\infty} h^2(iT) \; = \; \frac{1}{12} \sum_{i=0}^{\infty} h^2(iT)$$

$$\doteqdot \; \frac{1}{12\,T} \int_0^{\infty} h^2(t) \, dt \quad \text{... provided } T \text{ is small enough.}$$

When T is much less than the time constants of the controlled plant, i.e. when sampling is rapid, the response $h(t)$ will be a function not of T but of the criteria used for setting the controller. In this case, the measured variable's variance due to output rounding is proportional to sampling rate $1/T$.

Thus the deadband and the dynamic errors due to rounding increase with sampling rate. This relationship has been observed both in the laboratory (see appended paper 'Performance of ...', Fig.5) and in an industrial setting. Its significance is to place upper bounds on sampling rate where normally only lower bounds are thought to apply. A cyclic controller's period has to be adjusted to suit the process under control; values of between 1/5 and 1/20 the process time constant are usually appropriate. In any particular case, the state into which a control system settles - static within a deadband or following some complex limit cycle - depends on the initial conditions.

Continuous controllers, and cyclic controllers that do not introduce cumulative output errors, are not subject to the phenomenon just described. However they too can cause cyclic behaviour of small apmlitude due to phase hysteresis and patterning errors from frequency scalers (especially the BRM). Such cycles are usually of small amplitude and high frequency. If the scalers in a continuous controller are coarse the controller output frequency may be subject to substantial step changes. If they are very fine (and therefore long), patterning errors will be larger and of longer duration. Unless designed to the contrary, continuous p.f.m. controllers cause hunting and a comprimise must be made between minimising errors and minimising actuator wear.

4.5.2 Stability

Control theorists have made a number of studies of stability in systems containing pulse frequency modulators. In every case they have considered closed loops comprising a modulator and a linear continuous process. A rather wide range of modulators have been examined, from simple integral pulse frequency modulators (IPFM), to ones combining frequency modulation with width[425] or amplitude modulation[207]. Pseudo-describing functions have been used (Lee & Jones [424], Pavlidis & Jury[204], Dymkov[205]); Lyapunov functions have been proposed (Blanchard[206], Clark & Noges[216], Shortle & Alexandro[208], Kun'sevitch & Chekhovoi[425]); Popov's method has been employed in a special case (Monopoli & Wylie[217]). These studies have been directed towards obtaining measures of the amplitude or existence of small limit cycles.

Limit cycles will certainly occur (save in the special case of deadband behaviour) if the controller, actuator or process contains an integrator (i.e. a pole at the origin). The size of these cycles depends on controller design, on the nature of the controlled process, on the initial conditions and especially on the coarseness of actuator quantisation. The studies cited above all relate to unrealistically simple systems and it is uncertain whether they could usefully be extended to practical ones containing coincidence gates or scalers. Experimental work has indicated that where pulse frequencies are a large multiple of the frequencies present in the modulating information, limit cycles cause very small deviations in the measured variable.

Instability-in-the-large can be approximately predicted from calculations wherein pulse-frequency operations are replaced by equivalent continuous (analogue) operations. While it seems likely that p.f. effects will reduce gain or phase margins, experimental evidence points to this reduction being slight - a few percent or a few degrees. Cyclic pulse-frequency controllers exhibit modified sampled data system behaviour; slow sampling reduces stability.

There is a need for further analysis of the conditions for instability in p.f.m. systems, preferably of a stochastic kind that would yield some statistics of limit cycles without requiring very exact system descriptions. Most useful of all would be an indication of the bounds within which a p.f.m. system can safely be analysed using traditional continuous system methods.

CHAPTER 5

CONCLUSIONS

The purpose of these studies has been to establish whether or not pulse frequency modulation and digital pulse techniques could be usefully combined for the control of processes. A general conclusion is that they can, and that the combination offers many of the advantages of digital control at a relatively low cost. However, the complexity of pulse circuits exceeds that of analogue circuits, so that the latter should only be replaced by the former in a number of special circumstances, such as when long term stability or high resolution is required. For every arithmetic or differential operation investigated, a pulse technique has been found or developed. For every variable considered, a pulse frequency modulating transducer of acceptable complexity has been described. Two process controllers of different designs have been tested and found to offer dynamic responses comparable with those obtained with analogue controllers.

It is suggested that the particular role of the pulse frequency controller is that of a semi-autonomous unit in a multi-loop or multi-level control scheme. Such units are required to mediate between the complex activities of a computer or human operator and the simple activities of transducers and actuators. Pulse frequency modulation is an excellent language for such mediation, while the presence of many units working simultaneously, and in parallel, removes the need for very high speed operations.

The thesis has been divided into three main parts, viz. descriptions respectively of pulse frequency modulation, digital pulse techniques and control applications. Detailed findings in these three parts are as follows.

Pulse frequency modulation is suitable for the transmission of high accuracy signals over fading channels. Where the channel has a very wide bandwidth and is free from phase distortion, p.f.m. has excellent noise rejecting properties. Where the channel bandwidth is very small, a p.f.m. signal is effectively demodulated during transmission and arrives as an attenuated version of the modulating signal. Between these extremes a wide range of demodulation circuits can be employed to give varying degrees of noise rejection. For process control using electrical signals a high ratio of channel bandwidth to information bandwidth is usually available: pulse frequency modulation puts this ratio to good use with the aid of fairly simple yet very linear circuits. Compared with more commonly used signal forms, p.f.m. is particularly suited to telemetry and unsuited to speech transmission; this partly explains its limited application in the past. P.f.m. is wasteful of bandwidth, and consequently not well suited for multiplexing.

Circuits for manipulating the information carried in the frequency of a pulse train can usually be assembled from four basic elements. These are the anti-coincidence gate, the bidirectional counter, the frequency scaler and the pulse canceller, for all of which designs are given. Semi-synchronous counters appear most suited to pulse work, having adequate speed, short transition times and being simple to cascade. Two contenders for the frequency scaling function are the binary rate multiplier (BRM) and binary rate divider (BRD). The former is quite widely used and employs the technique of pulse frequency synthesis. The latter has not apparently been used for frequency scaling, it makes use of counter overflow (following addition) in the manner of a direct digital analyser unit. By ~~each of~~ four error criteria, each appropriate to a different application, the BRD is superior to the BRM.

Frequency addition and subtraction require an anti-coincidence gate. Frequency integration is implemented by a counter. Frequency differentiation is more complex, and all four methods described need quite elaborate circuits. Four methods of frequency to number conversion are also described, respectively: counting, frequency tracking, frequency comparison and period tracking. For low pulse-frequencies satisfying certain constraints, period tracking (using a novel circuit) is the best method of both frequency to number conversion and of frequency differentiation. The rounding errors in frequency to number conversion are a function of both frequency and of phase, these errors are greater than those occurring during analogue to digital conversion.

Techniques for computing with pulse frequencies are generally similar to those used with pulse number information. However non-linear functions of pulse number are easier to form than the same functions of pulse frequency. A problem common to both these signal forms is the handling of negative quantities; two sign notations have been shown to complement each other in their areas of usefulness. The manipulation of sign, and the realisation of such non linear operations as limiting, require special circuits beyond the four basic elements.

The application of pulse frequency technology to measurement is readily achieved, and is to be recommended where a digital display is favoured or where transmission over some distance is required. The majority of present day transducers are designed for analogue instruments, although a certain number of frequency modulating transducers are also in use. This number is increasing, and with the trend towards higher accuracy and data logging a good case can be made for choosing frequency (or pulse frequency) as a standard form for measurement signals.

The extension from measurement to control introduces actuators which are not very amenable to direct operation by pulse rate signals. Control also requires memory, which can be achieved with some difficulty using analogue devices, with ease using digital devices and not at all using ephemeral frequency signals. The application of pulse frequency techniques to control entails parallel digital techniques for memory operations and (sometimes) analogue techniques for actuation. The intermediate position of pulse frequency between the analogue and the digital signal forms permits easy conversions to and from them.

The two process controllers studied in detail both gave proportional and integral action. One was a cyclic device (Serck controller). The other was a novel continuous device. The former contained fewer circuit elements. The latter performed more functions (such as display) and was less susceptible to quantisation effects. It was found that in the particular case of a controller whose output feeds a digital integrator (e.g. a stepping motor), there is an unfortunate inter-reaction between sampling and rounding. At very high relative sampling rates, the variance of the controlled variable attributable to quantisation is roughly proportional to sampling rate. This phenomenon, a type of 'random walk', also appears in other digital calculations, for example simulation. The need to avoid over rapid sampling is an annoying constraint upon a cyclic controller that has no counterpart in a continuous controller. Continuous pulse techniques seem generally rather simpler than cyclic ones, having no need for sequencing logic, and no special start-up state. They are also easier to monitor, which is an aid to maintenance.

Given the proper commercial development of the four basic elements, continuous pulse frequency devices could be readily assembled to handle process instrumentation and control. Such devices offer higher accuracy, lower drift and greater operator convenience than their analogue counterparts and could be readily incorporated into a secure computer control scheme. Pulse frequency methods are extendable to other types of control provided information bandwidths of greater than about 100 Hz are not required.

BIBLIOGRAPHY FOR CHAPTER 2

201 Martin, J.D., 'Signal processing and computation using pulse-rate techniques' Radio & Electron. Engr. 38, 6, Dec. 1969, pp. 329-344

202 Ross, A.E., 'Theoretical study of pulse-frequency modulation' Proc. i.r.e. 37, 11, Nov. 1949, pp. 1277-86

203 Richards, G.A., Ph.D. Thesis, Southampton 1968

204 Pavlidis, T., and Jury, E.I., 'Analysis of a new class of pulse-frequency modulated feedback control systems', IEEE. Trans. I AC-10, 1 Jan. 65, pp. 35-43 and AC-10, 2, pp. 211-4

205 Dymkov, V.I., 'Periodic states in pulse-frequency systems' Avtomatica i Telemekhanika, 28, 11, Nov. 1967, pp. 101-8 (Engl. translation)

206 Blanchard, J.G., 'Theory and application of pulse-frequency modulated systems' Ph.D. Thesis, Berkeley 1966

207 Tait, K.E., 'An analysis and evaluation of signal-dependent sampling in discrete-continuous feedback control systems' Int. J. Control, 4, 3, pp. 201-239 and 4, 4, pp. 297-324

208 Shortle, G.E., & Alexandro, F.J. 'Stability of a class of pulse-frequeucy modulated control systems' Trans. IEEE, AC-15, 4, Aug. 1970, pp. 508-10

209 Abate, J.E., 'linear and adaptive delta modulation' Proc. IEEE, 55, 3, March 1967, pp.298-308

210 Inose, H., Aoki, T., & Watanebe, K., 'Asynchronous delta-modulated system', Electron. Lett. 2, 3, March 1966, pp. 95-6

211 Sharma, P.D., 'Characteristics of asynchronous delta-modulation and binary slope quantised p.c.m. systems' Electron. Engg. 40, 479, Jan. 1968, pp. 32-7

212 Brolin, S.J., & Brown, J.M., 'Companded delta-modulation for telephony' IEEE Trans. COM-16, 1, Feb. 1968, pp. 157-62

213 Menshikov, G.G., 'Delta modulation with increased information' IFAC Pulse Symposium, Budapest, April 1968, paper F12 (in Russian)

214 Betts, J.A., & Ghani, N., 'Adaptive delta-modulation for telephony' Electron. Lett. 6, 11, pp. 336-8

215 Libois, L.J., 'A novel method of code modulation - delta modulation' L'Onde Electrique, 32, Jan. 1952, pp. 26-31 (English translation)

216 Clark, J.P.C. & Noges, E., 'The stability of pulse-frequency modulated closed loop control systems' IEEE Internat. Convent. Rec. 14, Pt. 6, 1966, pp. 179-85

217 Monopoli, R.V., & Wylie, B.K., 'A new model for stability analysis in systems with pulse-frequency modulators' IEEE Trans. AC-15, 3, June 1970, pp. 342-5

218 Čakulev, T.I., & Stojanović, B.M., 'High performance voltage-to-frequency converter' Electron. Engg. 41, 493, March 1969, pp. 346-8

219 Mackay, R.S., 'Bio-medical telemetry' (Sensing and transmitting biological information from animals and man), John Wiley, New York, 1968

220 Korytkowski, J., 'Accurate current-to-frequency converter' Electron. Engg., 39, 475, Sept. 1967, pp. 568-70

221 Tesic, S., 'Simple voltage-to-frequency converter with increased frequency range' Electron. Lett., 4, 2, Jan. 1968, pp. 28-9

222 Smith, K.C., & Sedra, A., 'Simple wideband linear voltage-to-frequency converter' Electron. Engg., 40, 481, March 1968, pp. 140-3

223 Marconero, R., & Pallotino, G.V., 'Wide-range current-to-frequency converter' Electron. Engg., 40, 483, May 1968, pp. 264-6

224 Bamford, R.J., 'A new current integrator' Trans. IEEE, IECI-14, 1, April 1967, pp. 22-7

225 Jury, E.I., & Blanchard, J.G., 'Non-linear discrete equivalence of IPFM systems', IEEE Trans., AC-12, 1967, pp 415-422.

226 Rowe, H.E., 'Signals and noise in communication systems' Van Norstrand, New York, 1965

227 Brown, J., & Glazier, E.V.D., 'Telecommunications' Chapman and Hall, London, 1969

228 O'Grady, W.P., 'A linear frequency-to-voltage converter' Electronic Engg., 36, 441, Nov. 1964, pp. 776-8

229 Brown, R.H., 'Low-frequency averaging and instantaneous ratemeters' Electronic Engg., 40, 479, Jan. 1968, pp. 2-8

230 Lathi, B.P., 'An introduction to random signals and communication theory' International Textbook, London, 1970

231 Helstrom, C.W., 'Statistical theory of signal detection', 2nd Ed., Pergamon, London, 1968

BIBLIOGRAPHY FOR CHAPTER 3

301 Lundh, Y. 'Digital techniques for small computations' 19, 1, (Jan. 1959) pp 37-44

302 Moshos, G. J., 'Design of real time computers utilising counting techniques' NASA Report TN D-3042 (1965)

303 Nicola, R. N., 'Operational digital techniques for special purpose computers' Aeronautical Engineering Review, March 1956.

304 Mayorov, F. V. and Chu, Y., 'Digital differential analysers', Iliffe Books, 1964

305 Leonhard, W., 'Zählende Rechenschaltungen für Regelaufgaben', Archiv. für Electrotechnik, 49, 4, 1964, pp 215-234 (In German)

306 Moshos, G. J., 'Survey of countup-countdown machines', IFAC Pulse Symposium, Budapest, April 1968, Paper A-23

307 Unger, S. H. 'Hazards and delays in asynchronous sequential switching circuits' IRE Trans. Circuit Theory, CT-6, March 1959, pp 12-15

308 Meisel, W. S., and Kashef, R. S., 'Hazards in asynchronous sequential circuits', IEEE Trans. Computers, C-18, 8, Aug. 1969, pp 752-759

309 Gruzdev, S. V., and Karpov, R. G., 'Ferrite transistor subtracting devices for pulse-frequency signals', Instrum. & Exp. Techn., Jan.-Feb. 1966, pp 108-112 in English translation.

310 Kemploy, J. R., and Vernon, P., 'An efficient method for using an audio magnetic tape recorder for storing random pulses', J. Sci. Insti., 44, 7, July 1967, pp 566-8

311 Vincent, C. H., 'The comparison between a special shift register and a scaler for derandomisation', J. Sci. Insti. Ser. 2, 1, 1, Jan. 1968, pp 22-4

312 Anon., 'Principles of control, counting techniques - Part 5', Mullard Ltd. Electronics and Control Review, June 1968, p 3

313 Gaines, B. R., and Joyce, P. L., 'Phase computers', Proc. 5th A.I.C.A. Congress, 1967, pp 48-57

314 Parsons, B., 'High speed synchronous reversible binary and binary coded decimal counters', Texas Instruments, Application report B.40, 1968

315 Arnstein, W., Mergler, H. W., and Singer, R., 'Digital linear interpolation and the binary rate multiplier', Control Engng. 11, 6, June 1964, pp 79-83

316 Yang, Hsi-zeng, 'Determination of maximum error of a binary multiplier', Automation & Remote Control, 21, 7, July 1960, pp 709-713 (pp 1007-1014 in original Russian edition)

317 Moshos, G. J., 'Error analysis of the binary rate multiplier' NASA report TN D-3124, 1965

318 Dunworth, A., and Roche, J. I., 'The error characteristics of the binary rate multiplier', IEEE Trans. on Computers, C-18, 8, Aug. 1969, pp 741-745

319 Butaev, G. M., and Romashkan, V. S., 'Linear code to pulse-frequency converter', Meas. Techn. No. 2, Feb. 1968, pp 194-196 (English translation)

320 Martin, J. D., 'Signal processing and computation using pulse-rate techniques', Radio & Electron. Eng'r. 38, 6, Dec. 1969, pp 329-344

321 Rosemount Engineering Company 'Diogenes' multi-loop controller. Patent.

322 Leonhard, W., and Müller, H., 'Stetig wirkender digitaler Drehzahlregler' ETZ-A, 83, 12, Dec. 1962, pp 381-387 (In German)

323 Talbot, G. C. A., and Senior, R., 'Rescaling a pulse train' Electronic Eng'ing. 40, 480, Feb. 1968, pp 95-96

324 Kaps, G., 'Scaling of frequency analogous measured values' IFAC Pulse symposium, Budapest, April 1968, Paper A-43

325 Kostka, P., 'Compteurs complémentaires', Automatisme 15, 1, Jan. 1970, pp 20-23 (In French)

326 Rey, T. J., 'Digital rate synthesis for frequency measurement and control', M.I.T., Lincoln Lab., Technical Report, No. 188, 30th Oct. 1958, ASTIA 205596

327 Wood, P., 'Frequency meter with continuous digital presentation', J. Brit. I.R.E., 26, 2, Aug. 1963, pp 109-13

328 Tiroshi, I., and Ben-Uri, J., 'Analysis of an up-down counter used as a velocity error servomechanism', IEEE Trans. AC-13, 1, Feb. 1968, pp 83-96

329 Levi, D., Alexandrovitz, A., and Ben-Uri, J., 'Improved Dynamic characteristics of reversible counter used as a velocity-servo error register through reduction of noise effect', Proc. I.E.E., 117, 4, April 1970, pp 831-4

330 Martin, J. D., 'Pulse-rate to digital code-tracking converters', Joint Conf. on Digital Methods of Measurement, Canterbury, 1969, Published as IERE Conference Proceedings No. 15, pp 95-104

331 Ladányi, L., 'A complex controller using digital-operational techniques', IFAC Pulse Symposium, Budapest, April 1968, Paper C33

332 * Thomas, T. H., and Hughes, M. T. G., 'Performance of a digital two-term controller', Automatica 5, 1, Jan. 1969, pp 67-75

333 Widrow, B., 'A study of rough amplitude quantisation by means of Nyquist sampling theorem', Trans. I.R.E. on Circuit Theory, CT- , Dec. 1956, pp 266-276

334 Knowles, J. B., and Edwards, R., 'Autocorrelation function of the quantisation error process for a sinusoid', Electronics Letters, 4, 9, (3rd May 1968) pp 180-181

335 Bennett, W. R., 'Spectra of quantised signals', Bell System Tech. J., 27, July 1948, pp 446-472

336 Monroe, A. J., 'Digital processes for sampled-data systems', Wiley, New York, 1962, Chap. 21

337 Totty, R. E., and Clark, G. C., 'Reconstruction error in waveform transmission', IEEE Trans., Information Theory, IT-13, 2, April 1967, pp 336-8

338 Kellogg, W. C., 'Information rates in sampling and quantisation', IEEE Trans. Information Theory, IT-13, July 1967, pp 506-511

339 Nightingale, J. M., and Richards, G. A., 'Error analysis in binary rate modulation systems', IFAC Pulse Symposium, Budapest, April 1968, Paper A21

340 Knowles, J. B., and Edwards, R., 'Computational error effects in a direct digital control system', Automatica, 4, 1, May 1966, pp 7-29

341 Stephenson, J. M., 'Low noise tachometer generator', NRDC Licence No. 51137/69

342 Weinmann, A., 'Digitale Systeme zur Messung, Regelung oder Begrenzung der Beschleunigung', Electrotechnik u Maschinbau, 81, 2, 15th Jan. 1964, pp 30-34

343 Martin, J. D., and Shiner, A. J., 'Pulse-rate exponential function generator', Electronics Lett. 7, 2, 28th Jan. 1971, pp 55-56

344 Winston, G. C., 'Curvilinear rate-multiplier for precise interpolation', Control Eng'g. 15, 4, April 1968, pp 81-82

345 Elsden, C. M., and Ley, A. J., 'A digital transfer-function analyser based on pulse-rate techniques', IFAC Pulse Symposium, Budapest, April 1968, Paper A41

346 Cogman, J. C., 'A dynamic analysis of a pulse-rate computing technique', Ph.D. Thesis, Southampton 1970

347 Vorobkevich, V. Yu., and Rakov, M. A., 'Frequency doublers and dividers using switching phase detectors', Telecom. & Radio Eng. 1967, 8, pp 115-118

348 Margolin, A. M., 'Frequency multipliers for digital frequency measuring devices: design principles', Izmeritel'naya Tekhnika, 1, Jan. 1967, pp 52-55

349 Sinha, N. K., Szabados, B., & Di Cenzo, C. D., 'A new high-precision digital tachometer', Electron. Lett., 7, 8, April 1971, pp 174-176

350 Thomas, T. H., 'A rapid pulse-frequency to number converter',
* Electron. Lett., _7_, 12, 17th June 1971, pp 345-6

* These two papers are bound in at the end of the thesis.

401 Sokolov, L.N., 'Errors in transducers with a frequency output', Meas. Techn., 1966, 7, pp 955-8

402 Knorring, V.G., 'Mechanism of the operation of measurement transducers with a frequency output', Meas. Techn., 1966, 8, pp 1045-3

403 Katkov, V.P., et al, 'Precision converters of "angle-code" type employing frequency transducers of quantum magnetometers', Aut. & Rem. Control, 1962, 2, pp 256-261

404 Severov, A.P., 'Pulse excitation of vibrating wire transducers', Pribory Sist. Upravleniya, 1968,2, pp 9-11

405 Milokhin, N.T., 'Frequency transducers for automatic information processing systems', Energiya, 1968

406 Rozen, Yu.V., 'Russian standard (GOST) proposals on frequency electric signals', Pribory Sist. Upravleniya, 1968, 6, pp 8-10 (in Russian)

407 Agar,J., 'Frequency modulating transducers', Rad. & Electron. Engr., 38, 2, Aug. 1968, pp 89-97

408 Anon. 'A quartz thermometer', Instrum. Practice, 19, 10, Oct. 1965, pp 907-10; (describes Hewlett Packard model 2801A thermometer, q.v.)

409 Smith, W.L., & Spencer, W.J., 'Quartz crystal thermometer for measuring temperature deviations in the 10^{-3} °C to 10^{-6} °C range', Rev. Sci. Instr., 34, 3, March 1963, pp 268-70

410 Rodely, A.E., White, D.F., & Chanaud, R.C., 'A digital flowmeter without moving parts', A.S.M.E. Winter Meeting, Chicago, Dec. 1965

411 Samson Apparatebau, 'Mechanical square-rooting device for differential-pressure flowmeters', U.K.Patt. 1046421; see also Control, 11, 103, Jan 1967, p 31

412 Price, E.F., & Chapman, Margaret, 'A sensitive low range force current pressure transducer', Control, 11, 105, Mar 1967, p 128

413 Blanchard,J., & Ivison, J.M., 'A short-range telemetering system for the measurement of temperature in rotating joints', Instrum. Pract., 21, 12, Dec 1967, pp 1128-32

414 Lövborg, L., 'A linear temperature-to-frequency converter', J. Sci. Instrum., 42, 8, Aug 1965, pp 611-4

415 Redding, R.J., 'The barrier method of safety in hazardous areas', Instrum. Pract., 21, 9, Sept 1967, pp 831-4

416 Abbott, J.L., 'Requirements for intrinsically safe electrical equipment', IEEE Trans., IGA-3, 1 Jan 1967, pp 40-42

417 A.E.I. Ltd., Liecester, 'Runway visual rangefinder', Specification, 1970

418 Gaines, B.R., 'Varieties of computer - their applications and interrelationships', IFAC Pulse Symposium, Budapest, April 1968, Paper c 33

419 Frederikson, T.R., 'The closed-loop step motor - an ideal actuator for process control', IFAC Pulse Symposium, Budapest, April 1968, Paper c 35

420 Andrews, J.P., 'The stepping motor as a machine tool actuator', Electron. & Power, Nov. 1971 (to be published).

421 Dummermuth, E.H., 'The digital differential integrator for real-time computing', IEEE Proceedings of Conf. Int. sobre Systemas, Redes y Computadores, Mexico 1971, pp 60-64

422 Bell, C.A., & Ray, D.J., 'The design of a digital controller', Brit. Chem. Engng., May 1969. (Warren Spring Laboratory 'Discola' instrument, now licenced to 4 companies)

423 Curry, E.E., 'The analysis of roundoff and truncation errors in a hybrid control system', IEEE Trans AC-12, Oct 1967, pp 601-4

424 Lee, S.S., & Jones, R.V., 'Integral control with pulse frequency modulation', IFAC General Symposium No.2, 1963, Section on discrete and self-calibrating systems. (N.B. Lee S.S. = Li C.C.)

425 Kuntsevitch, V.M., & Chekhovoi, Yu.N., 'Fundamentals of non-linear control systems with pulse-frequency and pulse-width modulation', Automatica, 7, 1971, pp 73-81

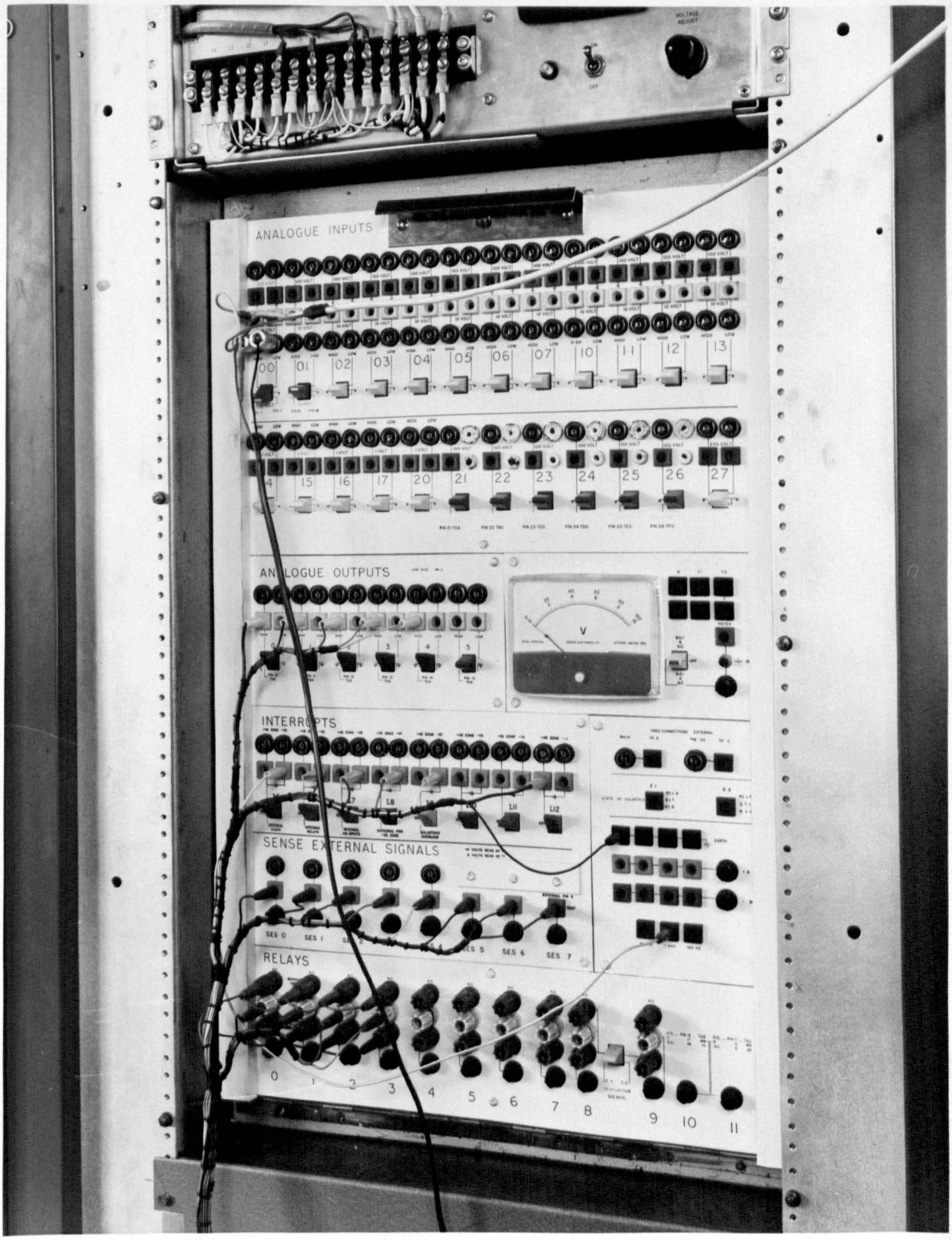


Photo C: General purpose computer interface.

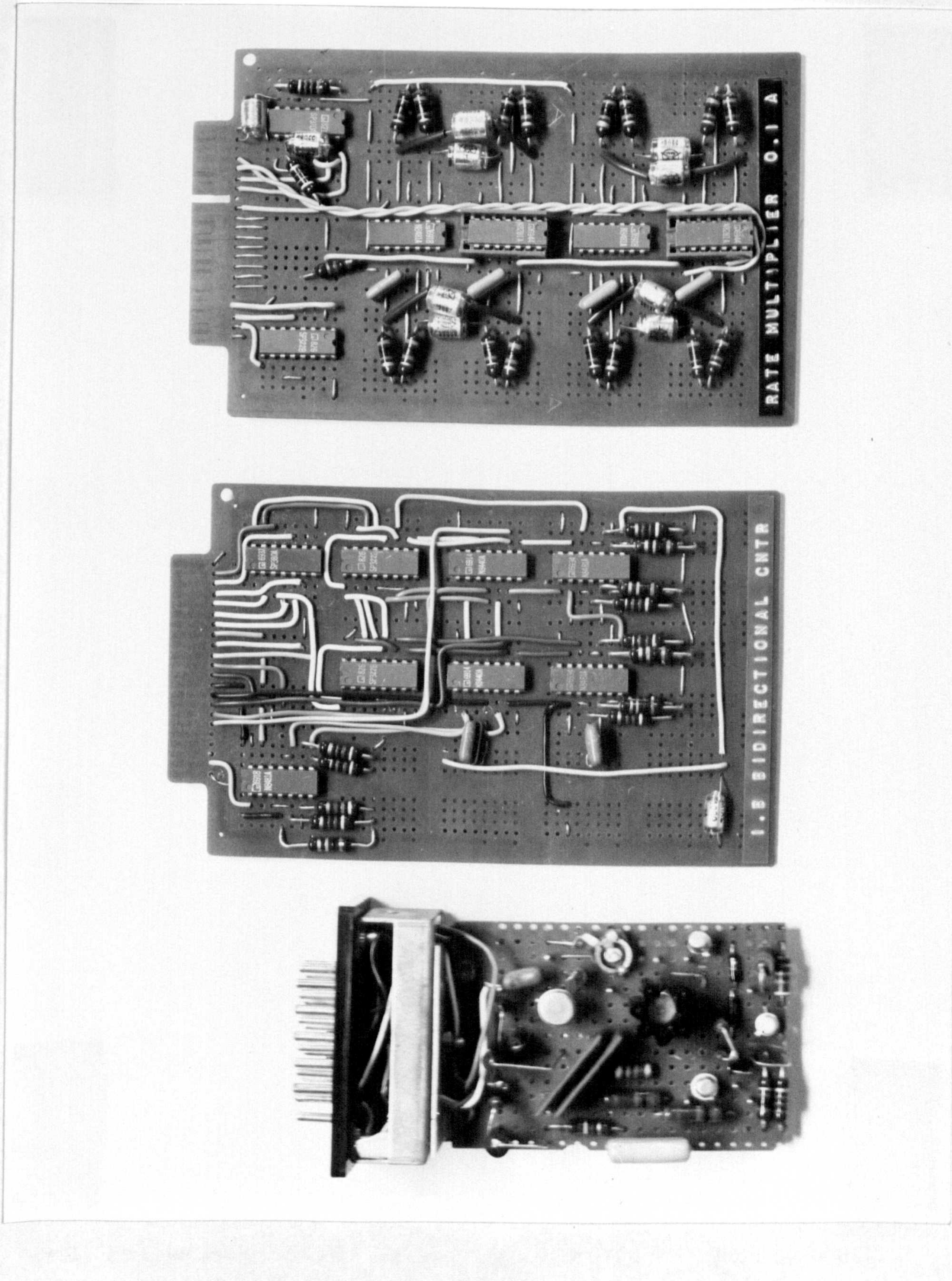


Photo D: Specimen circuits.

Appendix A: Circuits and wiring

Four sorts of circuits were constructed, namely a computer interface, transducer, actuator and digital circuits.

The interface was built as a general facility for all users of the School's GEC 90-2 digital computer. It handles analogue inputs and outputs, digital inputs and relay outputs, and interrupt signals. It provides circuits for protection, for testing, for scaling and for signal routing. A number of links to a Solartron 247 analogue computer are also incorporated, permitting simple hybrid computing to be undertaken. Part of the interface is shown in Photo C.

The transducer circuits have already been described.

The actuator circuits consist of relay drivers, inter-connected relays, contactors, solenoid valves, various pushbuttons and lamps. Parts of this wiring may be seen in Photo A.

The digital circuits were of prime interest. They were constructed with integrated circuit logic packages (TTL) of the dual-in-line type mounted on standard cards, as shown in Photo D, centre and right items. The 'Dualine' edge connectors that were used were designed to accept miniature jack pins. This allowed all the back wiring to be made with patch cords instead of soldered wires, an arrangement that proved very useful during circuit development.

The digital circuits were sited close to the actuator circuits in order to observe the effects of interference between the latter and the former. The a.c. relays in the actuator circuits were a prolific source of impulsive electrical interference. Voltage spikes considerably in excess of the logic circuits 'noise immunity' (= 1 volt) were induced in the inputs, outputs and power rails of the digital circuits. The most troublesome pick-up was that on the power rails, despite capacitive decoupling on every card. Addition of 0.01 F capacitors across the power rails of each rack annulled the principal interference effects. Standard '5 volt' integrated circuit logic is barely suitable for a noisy electrical environment; unfortunately, logic elements working off higher potentials are expensive, and are only available for simple logic functions.

The vulnerable points, in circuits that manipulated pulse trains, are the registers or counters, where signals are carried in parallel coded form. A single pulse on either power rail, on the clock line, on the set or reset line, or on the direction of count line, may totally change the value stored in a counter. Interference on the more significant of the parallel output lines can also cause large changes. By contrast, those parts of a circuit that handle signals in pulse-number or pulse frequency form are fairly resistant to gross errors: only the odd extra pulse is observed.

Digital controllers are much more vulnerable to electrical interference than analogue controllers, despite the superior noise immunity of PFM signals. Such controllers therefore need to be well screened, and to have appropriate filters on all lines entering or leaving them. All outputs from registers should be buffered. Earth links should be of large cross section. Each card should ideally have its own voltage regulator and supply line decoupler. Internally generated transients are generally small compared with those picked up from neighbouring higher powered equipment.

When the research was commenced, logic circuits were rather expensive, and medium scale integration (MSI) had not progressed far. It was clear from initial experiments that compact digital circuits of reasonable cost require MSI techniques, and that it would be useful to identify circuits that could be widely used as building blocks in digital systems. Too small a building block would require too many inter-block connections. Too complex a block would find few applications and too many distinct types would be required. A mounting card was chosen whose circuit capacity corresponded to that contained in about 10 basic logic packages, i.e. about 80 gates. The card allowed up to 20 external signal connections to be made. As a few MSI packages were employed (e.g. Quad Full Adder), the package count per card averaged about 7. Each card was intended to represent a future MSI package.

In order to increase their versatility, the building blocks developed initially were provided with many ancillary features. For example, the 4-bit bidirectional counter card requires 6 basic connections, but a further 12 were provided for refinements of the basic function.

As circuit design proceeded, a number of unforeseen operations were required and special circuits constructed. Circuit improvements were made. Limitations of cost and space justified the duplication of parts of the basic blocks, omitting the ancillary features. Eventually 50 cards were produced to 25 different designs. However the dominance of the original blocks is shown by the following figures:

Frequency scaler cards (4 or 8 bit)	11 made	
Bidirectional counter (4 bit)	7 made	Total 28
Anti-coincidence gate with canceller	6 made	out of 50
Dual pulse cancellers/freq. comparators	4 made	

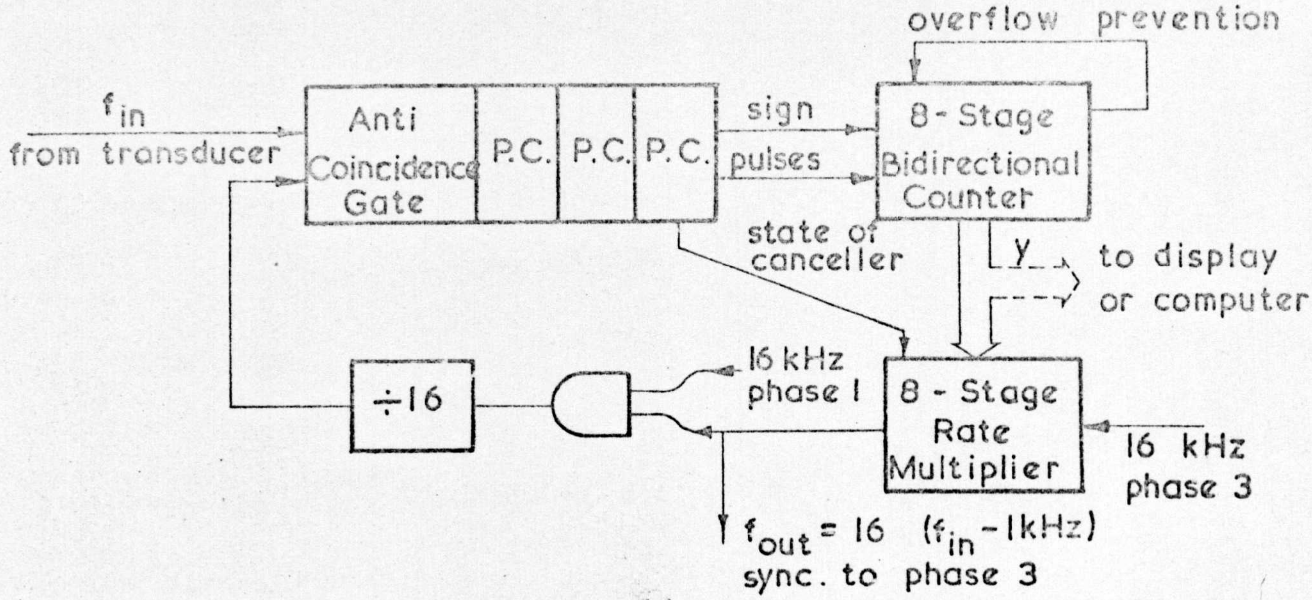
Integrated circuit bidirectional counters, and frequency scalers are now (1971) available as commercial MSI integrated circuits.

A configuration that often occurs in pulse circuits is a bidirectional counter connected in parallel to a frequency scaler. In its simplest form this would only require 5 external connections, regardless of length, and seems a good candidate for an MOST integrated circuit. Unfortunately the simplest form would be inadequate or clumsy for several common operations, for which parallel access to the counter is required.

The short list above omits two other circuits that are needed in almost every pulse manipulator, although any one device (e.g. controller) will usually require only one of each. These are a multiphase clock, and a pulse detector.

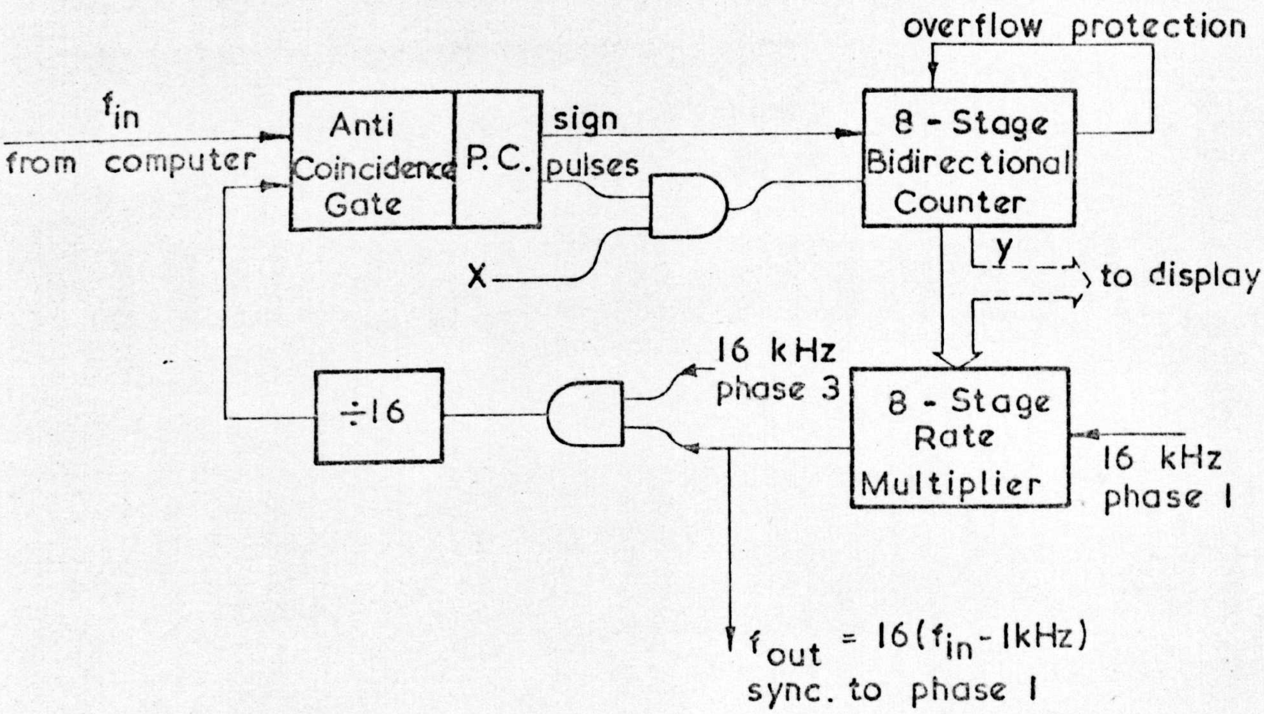
Figs. A1 to A6 show the block diagrams of the different parts of the two term process controller that was built. Fig. A7 shows the interconnections of the cards. Each letter indicates a different card type: the principal types are

A & R	Frequency scaler
B	Bidirectional counter
C	Anticoincidence gate
F & T	Clocks
L & M	Driver and receiver
Q,S & W	Cancellers/comparators
U & V	Routing, merging and dividing circuits
X,Z & AA	Logic, testing, meter drive, overlap prevention etc.



P.C. ≡ Pulse canceller

FIG. A1 Circuit to track without jitter and multiply a frequency.



Note: X is obtained from a frequency comparator; X='0' if $f_{in} < 1kHz$
If X='0', y and f_{out} are frozen in value.

FIG. A2 Circuit to track, multiply and preserve a remotely generated frequency.

A C G = Anti - coincidence gate

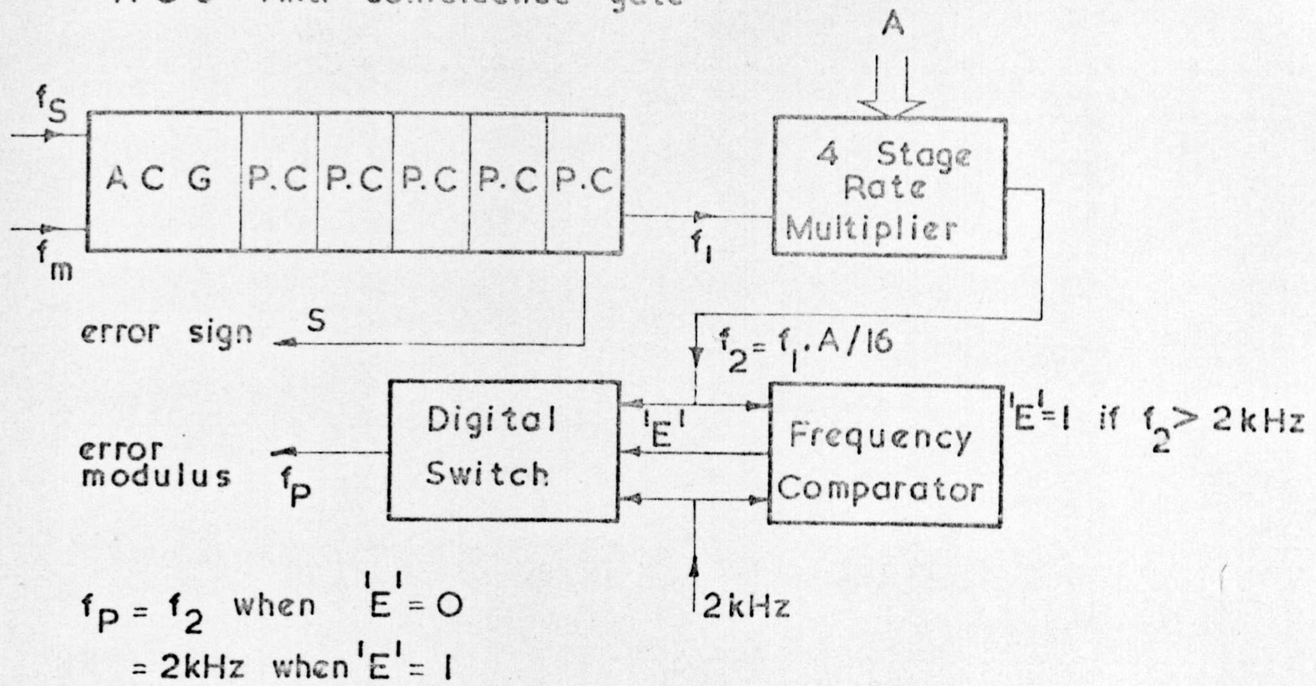


FIG A3 Circuit to generate and limit frequency f_P proportional to error

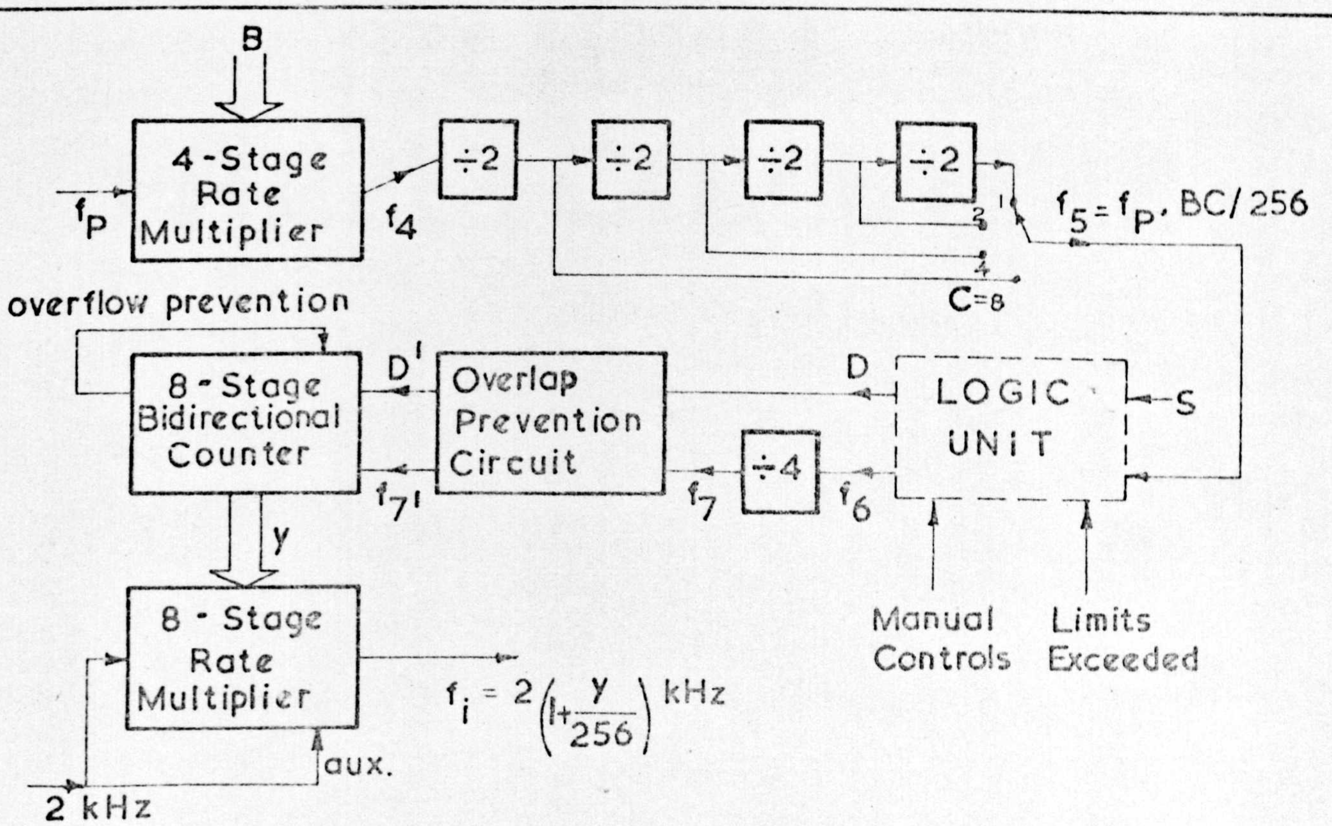
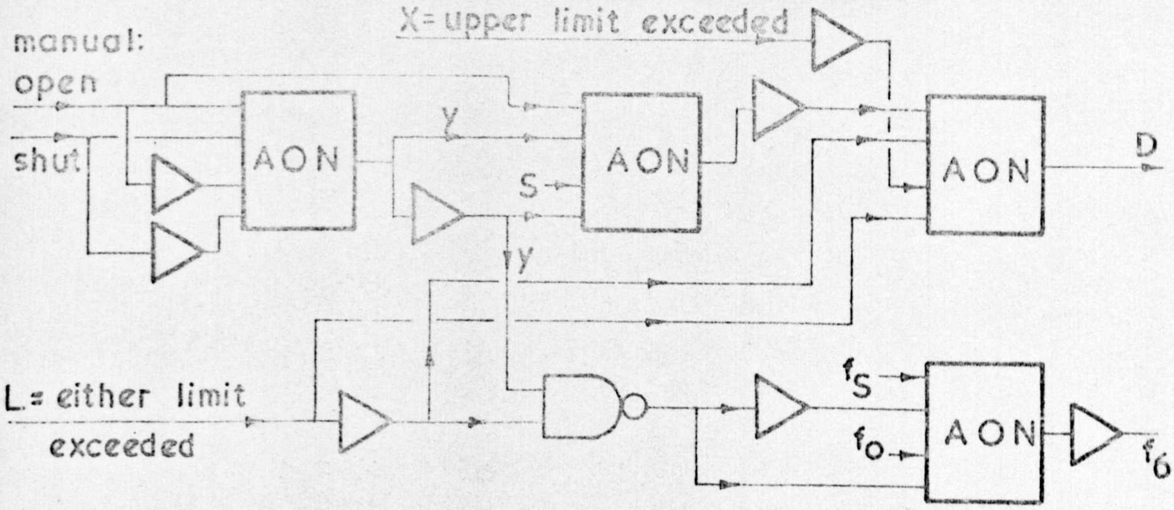


FIG A 4 Circuit to generate frequency f_i proportional to time--integral of error. T = 131/BC seconds

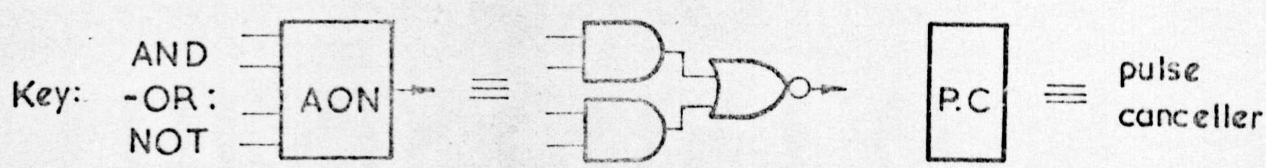


Y = open. $\overline{\text{shut}}$ + $\overline{\text{open}}$. shut, indicates one manual control operated

Sign out : $D = L\overline{X} + \overline{L}\,(Y.\,\overline{\text{open}} + \overline{Y}.S)$

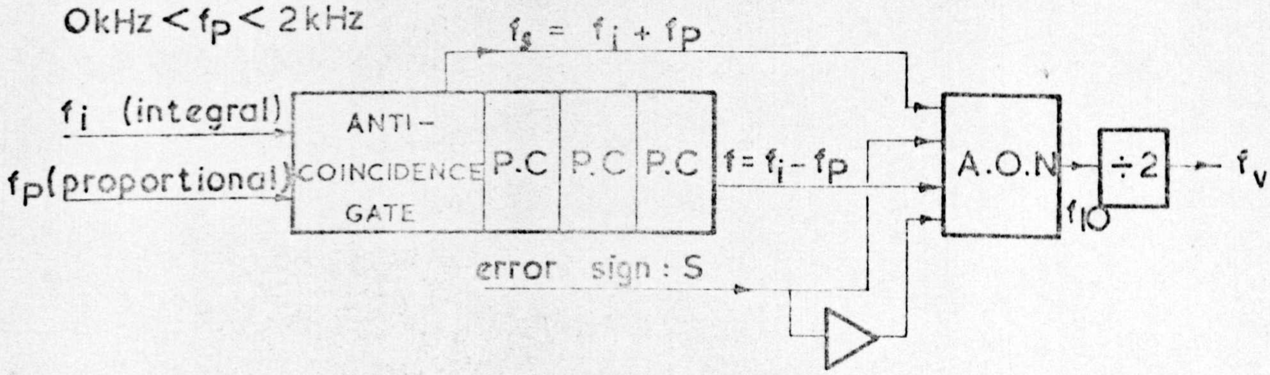
Frequency out: $f_6 = f_5$ if $L.Y = 0$, $f_6 = f_0$ if $L+Y=1$

FIG A5 Logic Unit



$2\text{kHz} < f_i < 4\text{kHz}$

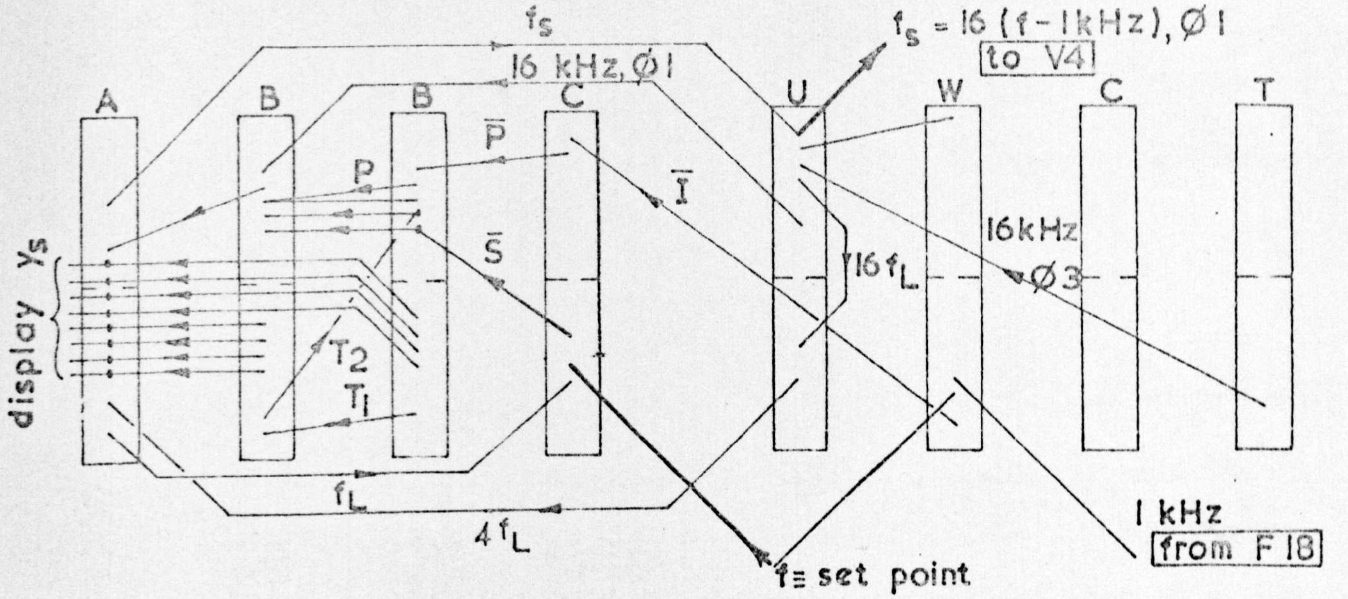
$0\text{kHz} < f_P < 2\text{kHz}$



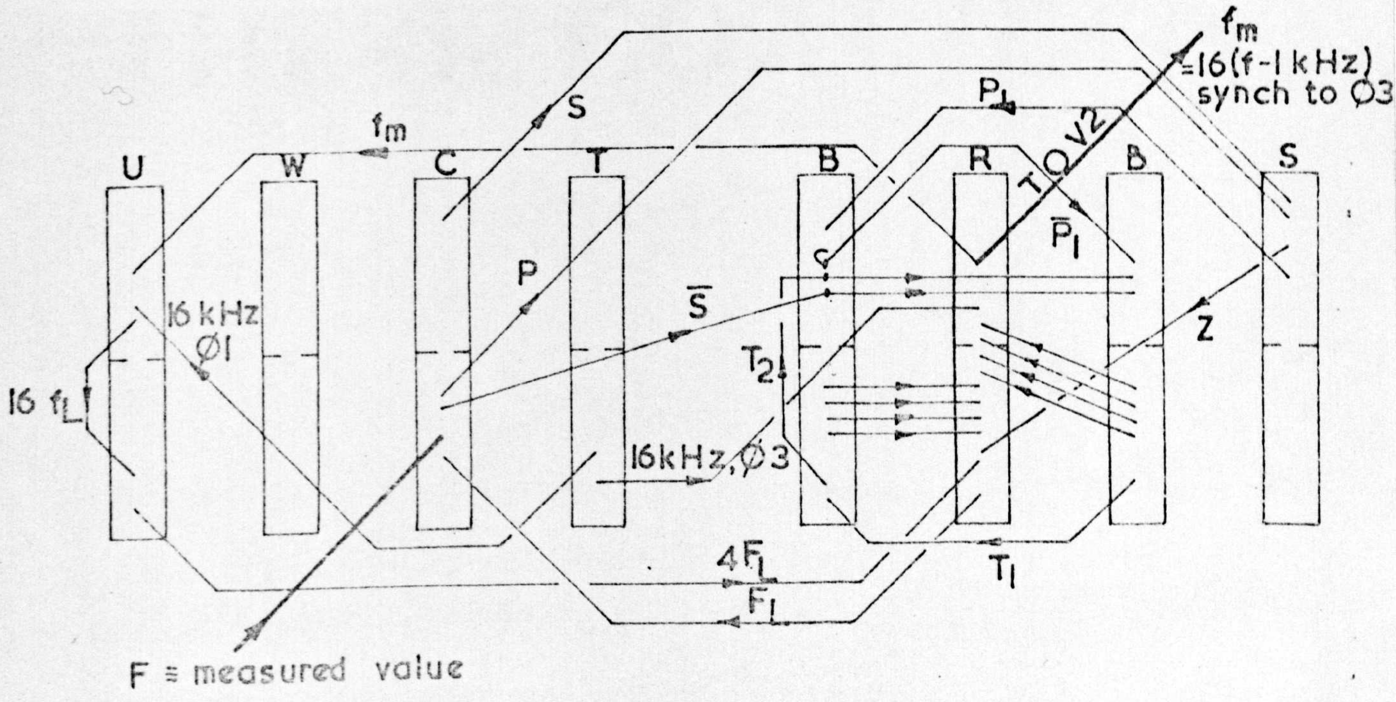
error + : $S = 1$, $f_V = f_s/2$

error − : $S = 0$, $f_V = f/2$

FIG A6 Combination of proportional and integral actions.



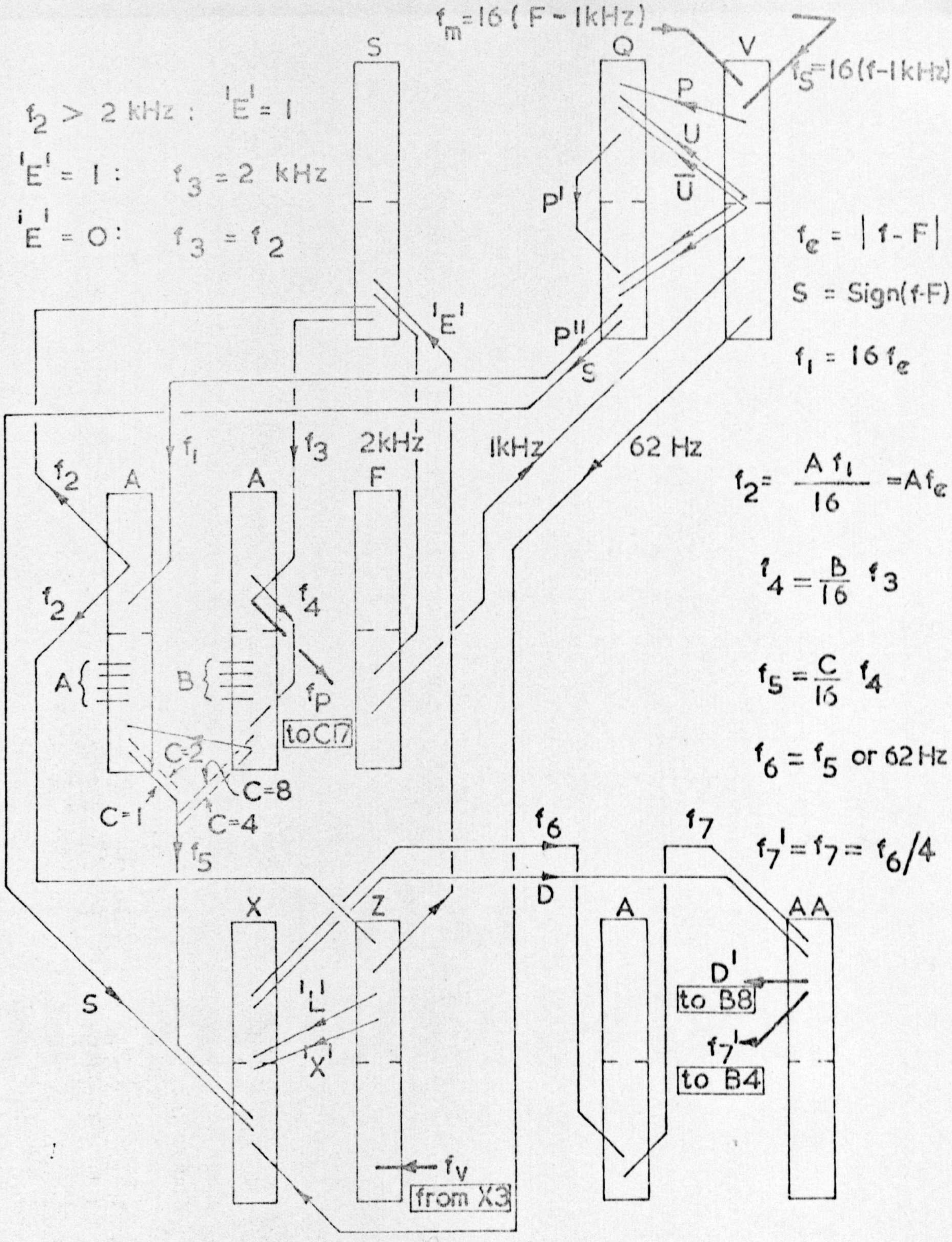
1. Tracking the set point



2. Tracking the measured value

FIG A7

CONNECTION DIAGRAMS FOR CONTROLLERS



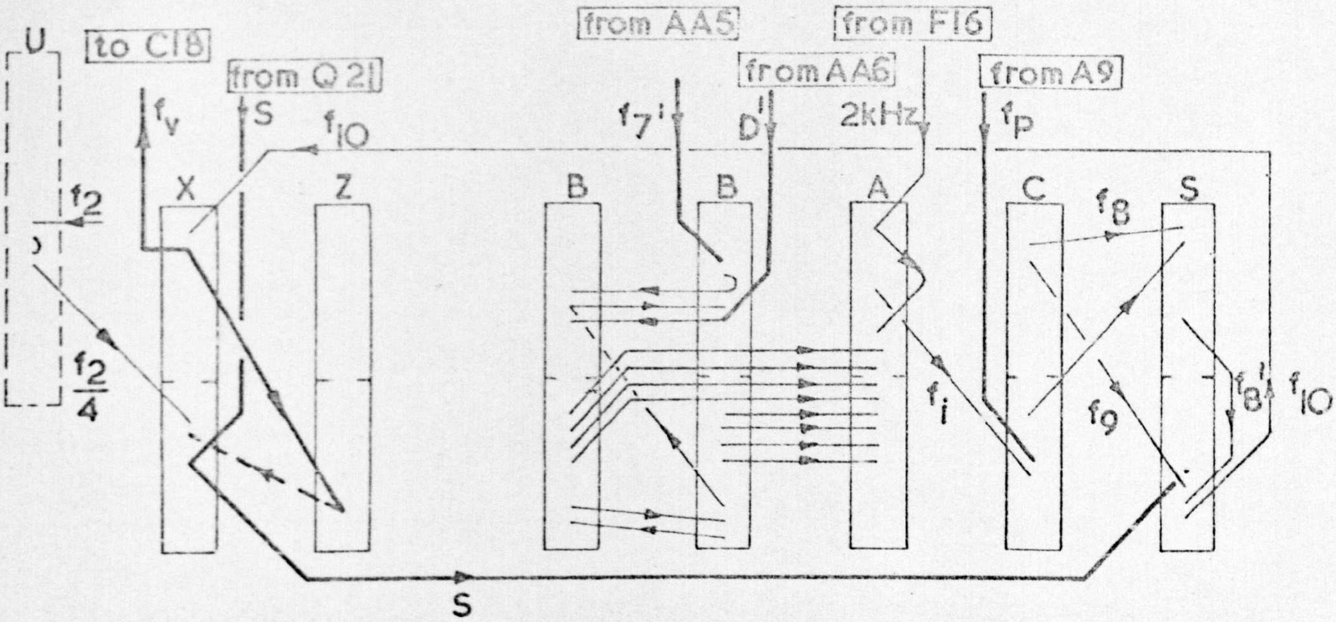
$'L' = 1$ if $f_v < 1\text{kHz}$ or $f_v > 2\text{kHz}$; $'X' = 1$ if $f_v > 2\text{kHz}$

For normal operation $f_P = A\ f_e$; $f_7' = B.C.\,f_P\ /1024$

3 Proportional and integrator - drive - frequency circuits

FIG A7

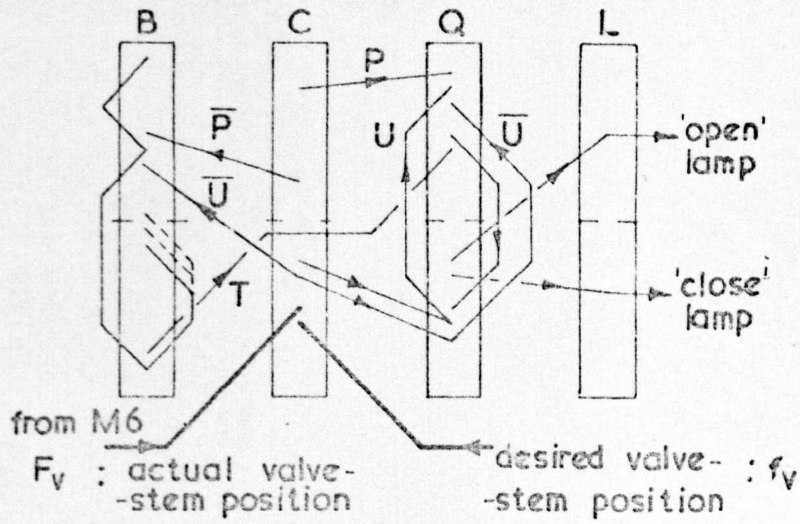
CONNECTION DIAGRAMS (cont.)



$f_8' = f_P \sim f_2 \qquad f_9 = f_P + f_i \qquad f_{10} = f_8' \text{ or } f_9$

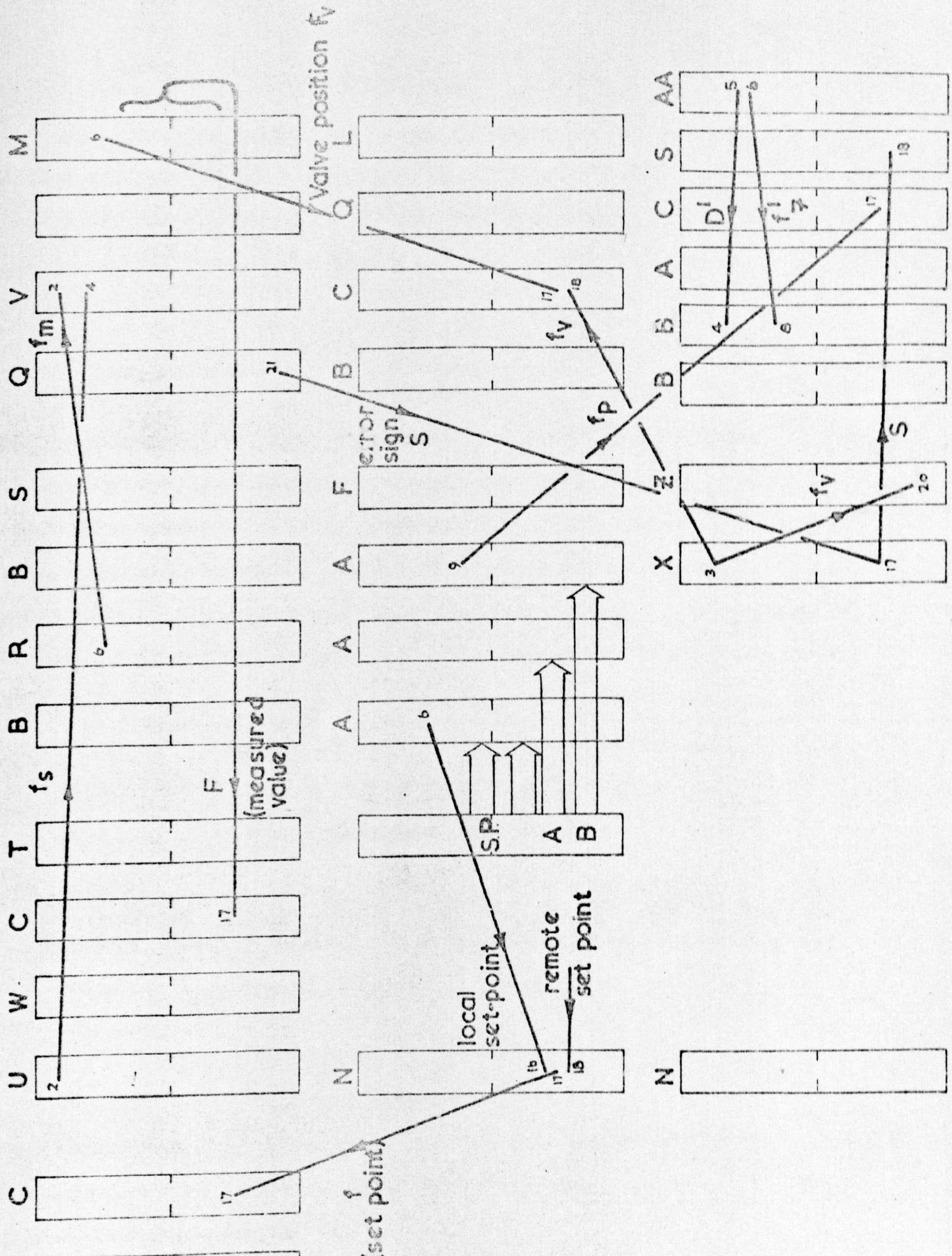
f_v = desired value-stem position = $f_{10}/2$

4. Output stage of proportional + integral controller



5. Valve-stem positioner circuit.

FIG A7
CONNECTION DIAGRAMS (cont)



6. Overall Layout

FIG A7 CONNECTION DIAGRAMS (cont.)

Appendix B: Computer Programming

Modelling was undertaken using two computers and two programming languages. For control studies a GEC 90-2 machine was used, while numerical analysis was performed in the University's Elliott 4130. In one particular case data tapes were prepared by the former for analysis by the latter.

The GEC 90-2 is a small American (SDS) computer designed for 'real time' applications. The School of Engineering's machine has 8192 words of 12 bit core store, and peripherals for handling paper tape. It has no backing store. Although a limited Fortran compiler is available, it is quite unsuited to control studies. The programmes for this computer were therefore written in its mnemonic assembly code 'Symbol'. During the period of the studies for this thesis considerable development of the 90-2 facility was undertaken. Interface construction has already been mentioned. General purpose programme development included the writing of the following 'library' routines and sub-routines:

DBLMUL	Signed double length multiplication
DUMPMOD	A compact non-relocatable dump
QUANTISE & ROUNDOWN	For truncating data
TYPEIN	A primitive 'conversation' organiser
DVLAG	A cyclic store
SIN 6 & SIN 10	Coarse high speed generators of sinusoids
DIAGINT	Interrupt diagnostic

The Elliott 4130 is a large computer organised for batch processing, and supplied with a wide range of input and output devices. Programmes for this machine were written in Algol.

Individual 'Symbol' programmes for the GEC 90-2

1. Title: Controller 4B 600 words 6 subroutines

Purpose: To model the Serck controller (of Chapter 4.3) in such a way as to allow 5 parameters, a set point profile, and process generated noise, to be specified.

Comments: The continuous counting operations of the Serck controller are represented by parallel digital operations in the computer: one cycle of the controller is represented by 12 time samples in the model. This model was used to investigate performance of the Serck controller with simple processes.

2. Title: Process 4B 150 words 3 subroutines

Purpose: To model a continuous process and act as a subroutine to the programme above. The process comprises a transport delay and a first order lag, each of which can be varied.

3. Title: Controller 5 500 words 5 subroutines

Purpose: To model the Serck controller in such a way as to permit the process under control to be simulated with an analogue computer (Solartron 247).

Comment: As the transport delay of a process is not amenable to analogue computer modelling, it is handled in the digital computer by a part of this programme. The basic cycle time of 0.36 seconds is broken into 18 parts to simulate controller continuity. Inadequate resolution in the digital-to-analogue converters led to the use of two in parallel, with digital separation and analogue recombination.

4. Title: BRM 47 words + 1024 workspace

Purpose: To generate on paper tape a sequence of characters representing the output of a 10 bit binary rate multiplier fed with 1024 pulses. The scaler factor is specified prior to each run.

Comment: The paper tape produced is used (to calculate power spectra) in the Elliott 4130 computer ... see under 'spectrum'.

5. Title: BRD 42 words + 1024 workspace

Purpose: As for 'BRM' but with a binary rate divider as the scaler.

6. Title: CRTLDS 750 words

Purpose: To employ the computer as a supervisor and monitor to a process controller. One frequency (a set point signal) is continuously output, five frequencies are input to the computer. Inputs are listed every 1/5 minute. A new output value is read from paper tape every minute.

Comments: All frequencies are in the range 1kHz to 2kHz. The output pulse train is generated by a digital to analogue converter that is updated every 100 microseconds. The output routine, effectively a binary rate divider, is very compact and fast.
The input signals use interrupts. 8 interrupts are employed to interleave four types of input/output operation.

7. Title: Staircase errors with binary rate multiplier 550 words

Purpose: To compare contents of a counter fed via a BRM with a uniform ramp. Obtain statistics of the deviation for each input pulse and each BRM factor from 1 to $2^N - 1$. There is also provision for inserting a divider between the BRM and the counter. The BRM length, N can be varied.

Specimen printout:

```
 1 IS DIVISION FACTOR
 7 STAGE COUNTER
1.555 IS MAX POS ERROR AT FACTOR 85 AT INPUT 85
1.555 IS MAX NEG ERROR AT FACTOR 85 AT INPUT 43
 .511 IS MAX STANDARD DEVIATION AT FACTOR 107
 .434 IS STANDARD DEVIATION OVER ALL FACTORS
```

8. Title: Staircase errors with binary rate divider 400 words

Purpose: BRD equivalent of programme above.

9. Title: BRM as pulse position modulator 400-500 words

Purpose: To treat the output of a binary rate multiplier as a position modulated pulse train. Calculates the statistics of this 'modulation', i.e. of position errors.

Comments: Various forms of this programme were used to examine errors as a function of stage in a cycle, of scaling factor and of counter length.

10. Title: BRD as pulse position modulator

Purpose: BRD equivalent of two forms of the programme above.

11. Title: Spectrum. (Uses procedures PLOTSP & FFT) 1050 words

Purpose: To plot the spectrum of a pulse train presented as a sequence on paper tape.

12. Title: PFM Spectrum (Uses PLOTSP & FFT) 1850 words

Purpose: To generate a pulse frequency modulated signal and then calculate its Fourier or power spectrum, which is then listed or plotted.

Comments: Each pulse generated is replaced by a truncated sin(x)/x waveform, which is then sampled at regular intervals. The truncation is chosen so as to give negligible alliasing errors during the Fourier transformation. Experiments were performed to establish the size of these errors. About 50 sidebands of the waveform were found to be necessary. The modulating signal was made either sinusoidal, or complex with a specified spectrum and approximate amplitude distribution.

13. Title: Detection (Uses NORMAL) 750 words

Purpose: To generate a uniform train of pulses, add Gaussian normal noise and apply the resultant waveform to a detector. The detector has filter, a variable threshold and is disabled for a period after each detection. The programme is for analysing detector performance.

Comments: Time is quantised into units of 1/10 times the pulse length. Due to excessive run times, the random noise was generated by a separate programme and stored in the disc file. Programme inputs are: interpulse period, digital filter coefficients, detector threshold, post detection inhibition time, presence or absence of signal clipping, noise level, number of pulses.

REPRINTED FROM

ELECTRONICS LETTERS, Vol.7, No.12, 17th June 1971, pp.345-6

RAPID PULSE-FREQUENCY–NUMBER CONVERTOR

Indexing terms: Analogue–digital conversion, Counting circuits

A circuit is described which will convert the repetition rate of a pulse train into a number, continuously but with very little lag. The circuit is considerably simpler than one recently described, and performs satisfactorily provided that the pulse rate does not vary too rapidly.

Introduction: In a recent letter, Sinha, Szabados and Di Cenzo[1] described a circuit to convert pulse frequency into a number. Their method was based on the arithmetic inversion of the measured interpulse period. They applied a bias to the frequency to be measured by adding a constant speed at the input of a speed-frequency encoder (i.e. a tachometer). This bias ensures that the rate at which speed is sampled does not fall below a specified minimum. The measurement of the period, rather than the rate, of pulses from a tachometer permits frequent and accurate estimates of speed to be obtained. These estimates are, however, very sensitive to any pulse jitter due to imperfect tachometer geometry. At high pulse rates, the accuracy is limited by the maximum speed of digital counters, although, in the particular context of speed measurement, this is rarely a problem.

Direct frequency-number conversion methods[2] have variously employed sampling, continuous tracking[3] and successive comparison.[4] The indirect method of conversion, via the measurement of interpulse period, can also be treated in these three ways. The method cited above[1] employs sampling and can accommodate rapid variations in frequency. However, this is achieved with considerable circuit complexity. For many applications, a much simpler circuit based on continuous tracking yields comparable results and is somewhat less demanding on tachometer geometric accuracy. This tracking circuit is described below.

Operation: The circuit shown in Fig. 1 essentially comprises two counters, a frequency scaler and a source of reference pulses of stable frequency f_r. Input pulses of the unknown frequency f control a series of gates in a closed loop: these are so connected that the frequency f_x of an internally produced pulse train tends to a fixed multiple of f. As f_x varies linearly with the numerical contents y of the second counter, y also tends to a fixed multiple of f, and frequency–number ($f \rightarrow y$) conversion is attained. Detailed operation is as follows.

Counter 1, initially zero, is fed with pulses at a rate f_x. On arrival of an input (e.g. tachometer) pulse, the f_x pulses are momentarily inhibited, the state of the most significant stage of the counter is preserved in bistable D, and the counter is then reset to zero. Simultaneously, one count is applied to counter 2 in a direction determined by the state of bistable D, D 'set' corresponding to a downwards count. Both counters are internally connected so as not to overflow. Since the input pulse serves four functions, denoted by R, I, F and C in the diagram, it must be suitably shaped to prevent race conditions occurring.

Two successive input pulses occur at times t_{i-1} and t_i. The retrospective interpulse period is then

$$p_i = t_i - t_{i-1} \text{ and } f(t_i) \simeq \frac{1}{p_i}$$

At instant t_i, the most significant bit of counter 1 will be set (neglecting phasing effects) provided that $p_i f_x \geq N$. If this is so, the arrival of the pulse at t_i will cause counter 2 to decrement by 1, which in turn will result in a fall in the pulse rate f_x. Alternatively, if $p_i f_x < N$, counter 2 will increment. When the input pulse rate is unvarying and equilibrium is reached,

$$f = \frac{1}{p} = \text{constant} \qquad pf_x = N \quad \ldots \quad (1)$$

neglecting those parts of the circuit shown dotted

$$f_x = f_r \frac{\bar{y}}{M} \quad \ldots \quad (2)$$

giving

$$\bar{y} = \frac{MN}{f_r} f = Cf \quad \ldots \quad (3)$$

The variables y and f_x in eqns. 1–3 are shown as averages. In practice, the circuit has a hysteresis of at least one bit, which results in a limit-cycle behaviour. This can be removed by the use of a pulse canceller, shown as broken lines in Fig. 1.

Accuracy: Provided that $N \geq M$, the resolution of the circuit is $1/M$. Several factors affect accuracy. The input pulse should be shaped to be very short, less than $1/f_r$. The reference-pulse train should be stable to better than $1/M$. Random phase relationships between the input and reference-pulse trains[5] cause slight uncertainty in eqn. 1.

The main cause of error is, however, the jitter introduced into the f_x pulses by the frequency-scaling circuit. The commonly used binary- and decimal-rate multipliers give substantial jitter,[6] although better designs exist.[7] The simplest way to reduce this jitter is to follow the scalar by a fixed-ratio integral frequency divider ($\div$K in Fig. 1), choosing f_r equal to $NK(f)_{max}$. In this way, it is possible to attain an accuracy of better than $2/M$, i.e.

$$y = \frac{MNK}{f_r} f + E = C' f + E \quad \ldots \quad (4)$$

where $|E| < 2$.

The addition of an auxiliary pulse train, as shown in Fig. 1, can offset bias in the incoming pulse frequency without loss of accuracy. The equation for equilibrium becomes

$$y = C'\left(f - \frac{f_b}{N}\right) + E \quad \ldots \quad (5)$$

The analysis above is in terms of equilibrium. Tracking that is fast enough to satisfy eqns. 4 or 5 is achieved, provided that

$$\left|\frac{dp}{dt}\right| = \frac{1}{f^2}\left|\frac{df}{dt}\right| < \frac{1}{N} \quad \ldots \quad (6)$$

Application: As eqn. 6 is not generally difficult to satisfy, the the circuit performs rapid frequency–number conversion. It can also be used as a frequency buffer and a frequency multiplier ($f_x = Nf$). The input to counter 2 is a 'signed' pulse train whose rate is a measure of the rate of change of f, and might represent acceleration. If the frequency-scaling circuit is replaced by a frequency divider[2] and other minor changes are made, a period-number tracking circuit is obtained.

T. H. THOMAS

Inter-University Institute of Engineering Control
School of Engineering Science
University of Warwick
Coventry CV4 7AL, England

24th May 1971

References

1 SINHA, N. K., SZABADOS, B., and DI CENZO, C. D.: 'New high-precision digital tachometer', *Electron. Lett.*, 1971, 7, pp. 174–176
2 MARTIN, J. D.: 'Signal processing and computation using pulse-rate techniques', *Radio Electron. Engr.*, 1969, 38, pp. 329–344
3 WOOD, P.: 'Frequency meter with continuous digital presentation', *J. Brit. Inst. Radio Engrs.*, 1963, **26**, pp. 109–113
4 MARTIN, J. D.: 'Pulse-rate to digital-code tracking converters'. IERE Conference Proceedings 15, London, 1969, pp. 95–104
5 THOMAS, T. H., and HUGHES, M. T. G.: 'Performance of a digital two-term controller', *Automatica*, 1969, **5**, pp. 67–75
6 MOSHOS, G. J.: 'Error analysis of binary rate multiplier'. US Government Report NASA : TN-D-3124, 1965
7 KAPS, G.: 'Scaling of frequency analogous measured value'. IFAC Pulse Symposium, Budapest, 1968, Paper A43

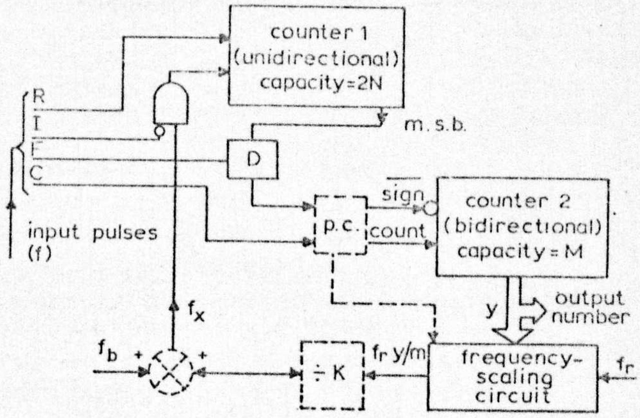


Fig. 1 *Frequency-tracking circuit*

p.c. is a pulse canceller, z is a single-bit variable which modifies the scaling factor from y to $y+1$. R = reset, I = inhibit, F = freeze, C = count, m.s.b. = most significant bit

Automatica, Vol. 5, pp. 67–75. Pergamon Press, 1969. Printed in Great Britain.

Performance of a Digital Two-Term Controller*

T. H. THOMAS† and M. T. G. HUGHES†

Digital techniques for process control are not limited to computer applications. A simple digital single-loop controller performs well if sampling rates and quantisation intervals are suitably chosen.

Summary—Digital techniques in process control are finding applications at several levels of complexity. In particular there are strong incentives to consider the use of small, single-loop digital controllers to replace their more conventional analogue counterparts. This paper presents the results of a combined theoretical and experimental study of a digital two-term controller, which operates upon the output of a variable-pulse-frequency measuring transducer, and generates at its output a pulse train to adjust a process variable via a stepping motor.

The distinguishing feature of such a system is that economic considerations demand relatively simple control algorithms. These introduce a number of non-linearities into the overall control equations. The paper considers the conditions under which these non-linearities and the effects of sampling are insignificant, so that the controller behaves effectively as a continuous device. Some of these non-linear effects are analysed, and it is shown that when they are significant they tend to degrade the performance of the controller. A parameter V, the ratio of dominant process time constant to sampling period, is used to define a region of satisfactory performance.

1. INTRODUCTION

FOR MANY years, the two-term‡ or three-term§ pneumatic analogue controller has been the preferred instrument for closed-loop process control. Electrical analogue controllers are now also widespread. Digital techniques permit a freedom from parameter drift that is advantageous in process work. However, the application of these techniques to process control has been largely limited to the use of relatively complex digital computers, whose high cost has been justified by the magnitude of the control task. In the process industries, there has been little opportunity to introduce digital control on a small scale before using it on a large scale.

The device under consideration in this paper is a single-loop digital controller, which uses pulse-rate techniques and can thereby be made comparable in cost with an analogue controller of similar performance. The controller is a member of a class of devices, sometimes known as countup-countdown machines [1], which employ pulse-frequency methods both for analogue-to-digital conversion and for arithmetic operations. Control action is based upon samples. Pulse-rate techniques are slow compared with the parallel transfer techniques used in the majority of computers. They are, however, more than fast enough for single-loop controllers. Set point and controller parameters are numbers in registers, and are thereby equally amenable to manual setting, local programming or alteration by a supervisory computer.

The equations associated with a controller using pulse-rate techniques are too complex for normal use in calculating settings, when such a controller is used in a plant. One convenient approach is to firstly define a near equivalent analogue controller and analyse its performance, and secondly consider the ways in which the performance of the actual digital device may be expected to differ. In this paper, such an approach is used to analyse one particular digital controller.

2. LIST OF SYMBOLS

Symbol	Purpose	Defined by:
A	Integral action parameter	(6)
a'	Normalised integral parameter for analogue equivalent controller	Fig. 3b
a	Integral parameter of conventional controller	Fig. 3c
B	"Proportional" action parameter	(10)
$c(t)$	Controlled (measured) variable	Fig. 1
$D(p)$	Controller transform	(27)
$e(t)$	Error in controlled variable	(15)
$f(t)$	Instantaneous frequency of transducer output	Fig. 1
F	Reference oscillator freq.	Fig. 1
$G(p), G'(p)$	Process transform	Fig. 3b

* Manuscript received 10 May 1968 and in revised form 16 September 1968. Recommended for possible publication by associate editor A. Sage; it was presented in a somewhat different version at the 1968 IFAC pulse symposium in Budapest, Hungary.

† University of Warwick, England.

‡ Proportional plus integral compensation.

§ Proportional plus integral plus derivative compensation.

Symbol	Purpose	Defined by:
H		(38)
k_1	Transducer conversion factor	(1)
k_2	A/D conversion factor	(13)
k_3	Offset	(4)
k_4	Process gain	(23)
L_1, L_2		(30)
m	Number in register	Fig. 2
p	Laplace transform variable	
q	Quantisation error	(5) etc.
$r(t)$	Desired output	(4)
S	Motor speed	
t, etc.	Time instants	Fig. 2
T	Sampling period	Fig. 2
u	Gating period	Fig. 2
V	Normalised sampling rate	(22)
W	Gain reduction factor	(29)
$x(t)$	Stepping motor position	Fig. 1
z		(15)
ε	Half dead-band	(32)
σ^2	Variance	
τ	Dominant process time constant	
$'$	notation used for per-unit quantities	
$<>$	Expected value	

3. REPRESENTATION OF THE CONTROLLER

3.1. *General description of operation*

Figure 1 illustrates application of the controller to the regulation of a variable, in this case temperature, $c(t)$. A transducer generates a pulse train of repetition frequency $f(t)$, where $f(t)$ varies linearly with $c(t)$. This frequency greatly exceeds that of the highest significant frequency component of $c(t)$, so that with negligible error the instantaneous value of $f(t)$ can be set equal to the reciprocal of the adjacent pulse spacing. A reference oscillator generates pulses of repetition rate F, both frequencies f and F being in the audio band. By counting pulses from the transducer, for a gating period defined by counting a fixed number of reference oscillator pulses, the ratio f/F is converted to a number, m, in a signal register. See Fig. 2. For F constant, as will be assumed throughout this paper, the number represents a short term running average of the controlled variable, $c(t)$.

Using the above method, $c(t)$ is converted to a number three times during each cycle of controller action. The first number, m_1, is subtracted from a number, m_r, representing a desired value. The resultant difference, m_a, represents error. The third number, m_3, is subtracted from the second,

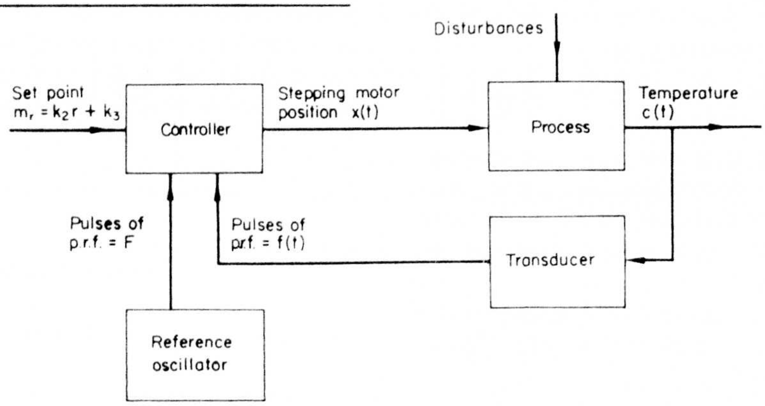


FIG. 1. Block diagram of control loop.

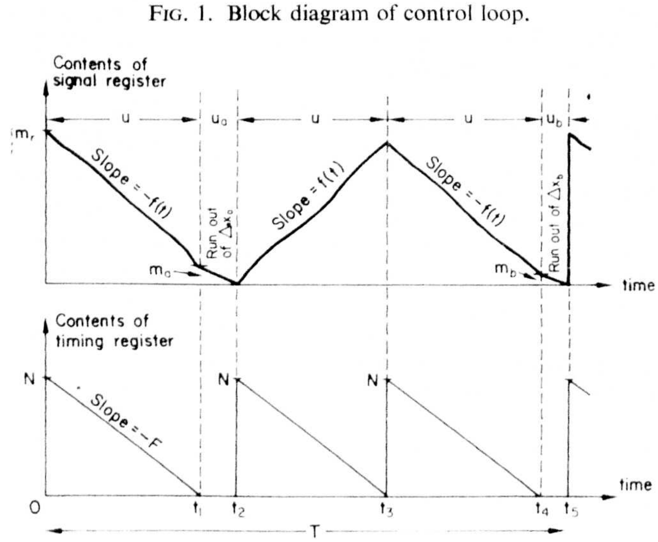


FIG. 2. Cycle of controller action.

m_2; the difference, m_b, is a measure of the rate of change of the controlled variable. The numbers m_a and m_b, multiplied by constants A and B respectively, are applied as step increments to a stepping motor. The multiplication and the transfer from register to motor are effected simultaneously by the use of rate multipliers and of an auxiliary pulse train derived from the reference oscillator.

The entire control cycle may thus be divided into four stages, as shown in Fig. 2.

Stage (i), $0 \leqslant t < t_1$, takes u seconds. $c(t)$ is converted to m_1 which is subtracted from m_r to leave m_a.

Stage (ii), $t_1 \leqslant t < t_2$, takes u_a seconds. The stepping motor is incremented by $A . m_a$ steps. u_a is usually very small.

Stage (iii), $t_2 \leqslant t < t_4$, takes $2u$ seconds. $c(t)$ is converted firstly to m_2, then to m_3. m_3 is subtracted from m_2 to leave m_b.

Stage (iv), $t_4 \leqslant t < t_5$, takes u_b seconds. The stepping motor is incremented by $B . m_b$ steps. u_b is usually very small.

The control action thus comprises application of motor increments proportional alternatively to error and to rate of change of controlled variable. Since the stepping motor acts as an integrator, this action is akin of that of an integral-plus-proportional $(P+I)$ analogue controller. The arithmetic operations are combined with the input and output operations.

3.2. *Full equations*

From the transducer

$$f(t)=k_1 c(t)+f_o \tag{1}$$

where k_1 and f_o are constants

The input conversion period, u, is defined by counting N references oscillator pulses, so

$$u=N/F\,. \tag{2}$$

During $0 \leqslant t < t_1$, transducer pulses are subtracted from the signal register. As initially the register contains setting m_r, at time $t=t_1$ the register contains

$$m_a=m_r-m_1 \tag{3}$$

where

$$m_r=k_2 r(t)+k_3 \tag{4}$$

and

$$m_1=\int_0^{t_1} f(t)\mathrm{d}t+q_1 \tag{5}$$

k_2 and k_3 are constants; $r(t)$ is the desired value; q_1 is round off error satisfying $-1<q_1<1$; and $t_1=u$.

The first stepping motor increment is an integral number of steps

$$\Delta_a x=A\,.\,m_a+q_a \tag{6}$$

where for A an integer, $q_a=0$; otherwise $-\tfrac{1}{2}< q_a \leqslant \tfrac{1}{2}$.

Time taken to effect the increment $\Delta_a x$ is

$$u_a=|\Delta_a x|/S \tag{7}$$

where S is stepping motor speed, and is constant. During the second and third conversion periods, transducer pulses are counted up in the signal register from time $t=t_2$ to $t=t_3$ and counted down from time $t=t_3$ to $t=t_4$. At $t=t_4$ the register contains

$$m_b=m_2-m_3 \tag{8}$$

where

$$m_2=\int_{t_2}^{t_3} f(t)\mathrm{d}t+q_2$$

and

$$m_3=\int_{t_3}^{t_4} f(t)\mathrm{d}t+q_3 \tag{9}$$

$t_3-t_2=t_4-t_3=u$; rounding errors satisfy $-1< q_2, q_3<1$.

The stepping motors second increment is

$$\Delta_b x=B\,.\,m_b+q_b \tag{10}$$

where for B an integer, $q_b=0$; otherwise $-\tfrac{1}{2}< q_b \leqslant \tfrac{1}{2}$.

The time to effect this increment is

$$u_b=|\Delta_b x|/S \tag{11}$$

giving a total cycle time

$$T=3u+u_a+u_b\,. \tag{12}$$

The constants in equations (1), (4) and (5) are inter-related so that when the value of the controlled variable equals the desired value, the error number m_a is zero. Hence

$$k_2=uk_1 \quad \text{and} \quad k_3=uf_0\,. \tag{13}$$

4. DEFINITION OF A NEAR EQUIVALENT ANALOGUE CONTROLLER

The equations, of the last section, themselves too numerous to give a useful insight to controller performance, are here simplified by the application of approximations.

Approximation A: neglect quantisation effects.

$$\text{assume } q_1=q_2=q_3=q_a=q_b=0 \tag{A}$$

Approximation B: neglect the small but variable part of the cycle time which is used to "run out" the stepping motor increments

$$\text{assume } u_a = u_b = 0, \qquad \text{hence } T = 3u\,. \tag{B}$$

Approximation C: neglect the attenuation of high frequency signal components caused by the method of analogue-to-digital conversion. Regard the number m_a in the signal register at time $t = t_1$ as the sample of a hypothetical continuous variable $m_a(t)$. Then provided $r(t)$ is maintained constant during the interval $0 < t < t_1$, $m_a(t)$ satisfies the transform relationship

$$\begin{aligned} M_a(p) &= k_1 \frac{(1 - e^{-pu})}{p} E(p) \\ &= k_1\, u e^{-p(u/2)} \left(\frac{e^{p(u/2)} - e^{-p(u/2)}}{pu} \right) E(p) \\ &= k_2 z^{-\frac{1}{6}} Tr(p) E(p) \end{aligned} \tag{14}$$

where

$$e(t) = r(t) - c(t)\,; \qquad z = e^{pT}$$

and

$$Tr(p) = \sin x / x \quad \text{for} \quad p = j\omega, \qquad x = \omega T/6\,. \tag{15}$$

For $\omega < \pi/T$, $Tr(p)$ satisfies $0{\cdot}95 < Tr(j\omega) \leqslant 1$; and as these are the frequencies of interest,

$$\text{assume } Tr(p) = 1 \tag{C}$$

(14) now simplifies to

$$M_a(p) \doteqdot k_2 z^{-\frac{1}{6}} E(p)\,. \tag{16}$$

Regarding m_b as a sample at $t = t_4$ of $m_b(t)$, then $m_b(t)$ must satisfy

$$M_b(p) = -k_1 u^2 p z^{-\frac{1}{3}} \,.\, Tr^2(p) \,.\, C(p) \tag{17}$$

and applying assumption (*C*) to (17), and using (13) and (*B*)

$$M_b(p) \doteqdot -k_1 u^2 p z^{-\frac{1}{3}} \,.\, C(p) = -k_2 (T/3) p z^{-\frac{1}{3}} \,.\, C(p) \tag{18}$$

A sampling controller satisfying equations (16) and (18) is shown in Fig. 3(a).

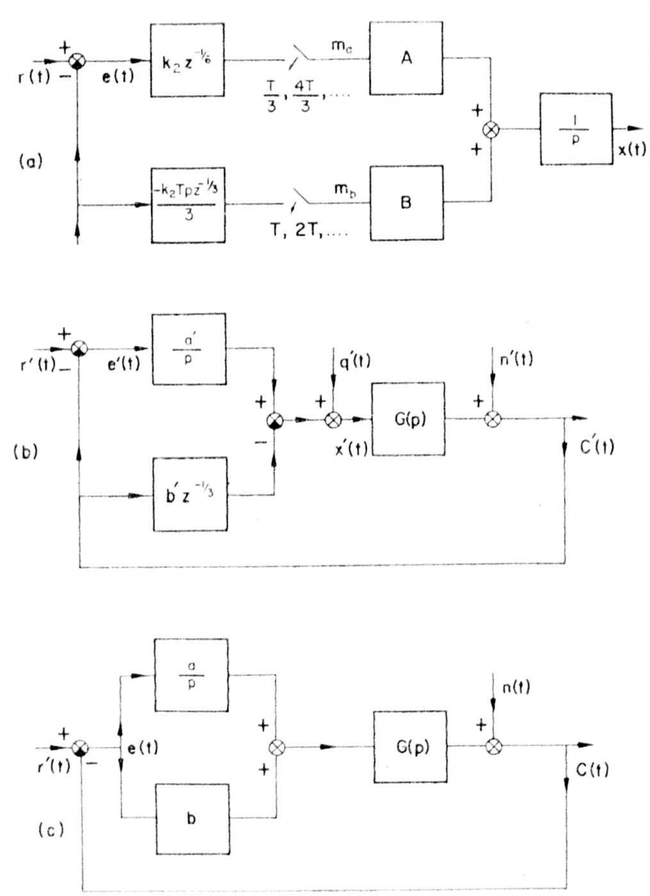


FIG. 3. (a) Controller simplified by assumptions (*A*), (*B*) and (*C*).
(b) Normalised control loop employing "near equivalent" analogue controller.
(c) Normalised control loop employing conventional $P + I$ controller.

Approximation D: neglect sampling effects. For negligible run-out times (assumption (*B*)), m_a and m_b may be regarded as impulse samples applied to an integrator (the stepping motor). So

$$X(p)+\frac{A}{p}M_a^*(p)+\frac{B}{p}M_b^*(p) \qquad (19)$$

and if the process characteristic heavily attenuates frequencies higher than $\frac{1}{2}T$, so that only lower frequencies are of interest

$$\text{assume } M_a^*(p)=\frac{M_a(p)}{T}, \text{ and } M_b^*(p)=\frac{M_b(p)}{T} \qquad (D)$$

giving from (19)

$$X(p)=\frac{Ak_2z^{-\frac{1}{3}}}{pT}.E(p)-\frac{Bk_2Tpz^{-\frac{1}{3}}}{pT.3}C(p)$$

$$=\frac{Ak_2}{T}.z^{-\frac{1}{3}}.\frac{E(p)}{p}-\frac{Bk_2}{3}z^{-\frac{1}{3}}C(p). \qquad (20)$$

For realistic processes and with (*D*) satisfied, $A \ll B$. Thus at high frequencies, for which the phase shift caused by factor $z^{-\frac{1}{3}}$ is significant, the first term of (20) will be very small compared with the second. Under these conditions, the factor $z^{-\frac{1}{3}}$ may be neglected, and

$$X(p)=\frac{Ak_2}{T}\frac{E(p)}{p}-\frac{Bk_2}{3}.z^{-\frac{1}{3}}C(p). \qquad (21)$$

5. PERFORMANCE OF A NEAR EQUIVALENT ANALOGUE CONTROLLER

Equation (21) expresses the action of an analogue approximation to the digital controller. If the dominant time constant of a controlled process is τ, then the sampling rate $1/T$ can be expressed in normalised form:

$$V=\tau/T. \qquad (22)$$

Process zero frequency gain is

$$k_4=\Delta c(t)/\Delta x(t). \qquad (23)$$

Then all variables can be re-expressed in "per unit" terms. By definition, per-unit process zero-frequency gain

$$\Delta c'(t)/\Delta x'(t)=1 \qquad (24)$$

using (′) as a per unit notation.
Writing

$$X'(p)=(a'/p)(R'(p)-C'(p))-b'z^{-\frac{1}{3}}C'(p) \qquad (25)$$

then

$$a'=Ak_2k_4/T\,; \qquad b'=Bk_2k_4/3\,;$$

$$z^{-\frac{1}{3}}=e^{-p\tau/3V}. \qquad (26)$$

Figure 3(b) shows the normalised control loop based on (25), while Fig. 3(c) shows for comparison a conventional $P+I$ controller, whose behaviour is well known.
For Fig. 3(b)

$$E'(p)=\frac{(1+b'z^{-\frac{1}{3}}G'(p))R'(p)-N'(p)}{1+D'(p)G'(p)} \qquad (27)$$

where

$$D'(p)=[(a'/p)+b'z^{-\frac{1}{3}}].$$

For Fig. 3(c), using undashed notation, and where $G(p)=G'(p)$

$$E(p)=\frac{R(p)-N(p)}{1+D(p)G(p)} \qquad (28)$$

where

$$D(p)=(a/p)+b.$$

The extra term $z^{-\frac{1}{3}}$ (i.e. a delay of $\tau/3V$) in the denominator of (27) will have significance only when V is small, say $V<6$. W is defined as the factor by which both a and b parameters must be reduced to restore optimum stability margins, following the introduction of a delay $\tau/3V$ as shown in (27)

$$\text{i.e.} \qquad W=\frac{a(\text{optimum})}{a'(\text{optimum})} \qquad (29)$$

Table 1 illustrates the effect of the delay for $V=3$. Note how, for such slow sampling, the permissible gains to control certain processes are much reduced.

TABLE 1. GAIN REDUCTION FACTOR, W, CALCULATED FOR VARIOUS PROCESSES

Process transfer function	W, for $V=3$
(i) $e^{-p\tau}/(1+p\tau)$	1·05
(ii) $e^{-0\cdot1p\tau}/1+p\tau)$	1·8
(iii) $1/(1+2p\tau+p^2\tau^2)$	2·6
(iv) $1/(1+20p\tau+p^2\tau^2)$	8·5
(v) $1/p(1+p\tau)$	2·0

The only effect of the factor $[1+b'z^{-\frac{1}{3}}G'(p)]$ in equation (27) is upon the response to set point changes. The step response of the controlled variable will in general display a lengthened rise time, increased integral of error squared and lower overshoot, due to this factor. The response of $c'(t)$ [Fig. 3(b)] to a set point step change is shown in Fig. 4 curve (a), for a typical process. Figure 4 curve (b) illustrates the response of $c'(t)$ to step noise $n'(t)$. This response is similar to that of $c(t)$ to set point change using a conventional controller.

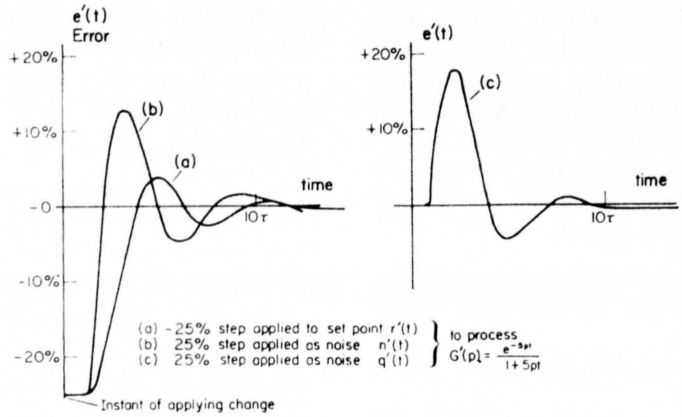


FIG. 4. Error responses of controller of Fig. 3(b).

6. SAMPLING EFFECTS

Approximation (C) was to neglect high frequency attenuation due to the method of analogue-to-digital conversion. Approximation (D) was to neglect intermodulation effects due to sampling. The attenuation and intermodulation effects partly cancel each other, and experiments show that only at fairly low relative sampling rates (say $V<5$) does performance differ appreciably from that predicted by using equation (21) to describe the controller. For $V<3$, however, control performance is sharply degraded, although control of some sort is usually feasible for $V<1$. For slow sampling, the frequency of any sustained oscillations tends to $\frac{1}{2}T$, at which frequency an impulsive sampler acts as a gain of $(1+e^{jd})/T$, where d may have any real value [2].

With this controller, performance is further slightly degraded by the alternate nature of the control action. The samplers of Fig. 3(a) do not close simultaneously. The z-transform relating error sequence $e^*(t)$ to the input sequence $r^*(t)$ (Fig. 3a) is that obtained by transforming equation (25) with additional higher order z terms in the denominator. These terms tend to zero as sampling rate is increased.

The numerical value of the control parameter A depends on sampling rate. For a given process and a given performance criterion, and assuming sampling is not too slow, optimum values of a' and b' may be calculated. For a typical process

$$a'_{\text{opt}}=L_1/\tau\,; \qquad b'_{\text{opt}}=L_2 \tag{30}$$

where L_1 and L_2 depend upon the type of process, and will have values of the order 1 and 10 respectively.

Substituting (26) in (30),

$$A_{\text{opt}}=\frac{TL_1}{k_2k_4\tau}=\frac{1}{V}\cdot\frac{L_1}{k_2k_4}\,; \qquad B_{\text{opt}}=\frac{3L_2}{k_2k_4}. \tag{31}$$

7. QUANTISATION

In this controller, the quantisation errors due to analogue-to-digital conversion, are represented by q_1, q_2 and q_3 in (5), and (9), while those due to arithmetic operations by q_a and q_b in (6) and (10). An exact analysis of quantisation errors is not profitable. MONROE [3] discusses two approaches to analysing quantisation effects. The first approach applies to a control loop in which variables are drifting slowly; it consists of determining a dead band within which no control action is taken. The second approach, due originally to WIDROW [4] ,applies to a control loop in which inputs to any quantisers range over several quantiser levels during each sampling period. Under these conditions the effect of quantisation on a signal is similar to the addition of white noise (of known power) to the signal.

Using the first approach for this controller, the dead band is, found to be a function only of the integral action. Defining dead band as $[c(t)-r(t)]=\pm\varepsilon$, then ε is that error in $c(t)$ which will just cause the integral action to increment the stepping motor by one step.

$$\therefore Am_a=Ak_2\varepsilon=1$$

giving dead band

$$\pm\varepsilon=\pm 1/k_2A \tag{32}$$

and assuming $A=A_{\text{opt}}$, then from (31)

$$\text{dead band}=\pm Vk_4/L_1\,. \tag{33}$$

The dead band increases with normalised sampling rate, V and process steady state gain, k_4. The latter can be reduced by gearing down (and unfortunately slowing down) the stepping motor.

The second, stochastic, approach to quantisation uses statistics developed in the appendix. The analogue-to-digital conversion errors q_1, q_2 and q_3 are partly intercorrelated, and each has a variance of $\frac{1}{6}$, while (q_2-q_3) has a variance of $\frac{1}{2}$. The arithmetic rounding errors q_a and q_b have variance $\frac{1}{12}$.

At the end of each cycle the stepping motor will have moved in error by q steps, where

$$q = Aq_1 + B(q_2 - q_3) + q_a + q_b. \qquad (34)$$

For relatively rapid sampling, for which quantisation effects are significant, $B \gg A$, (as $A \propto 1/V$). Neglecting the first term in (34), the error step during the ith cycle will be

$$q_i \doteqdot B(q_2 - q_3)_i + (q_a + q_b)_i$$

and variance of q_i for all i is

$$< q_i^2 > = B^2/2 + 2 \times \tfrac{1}{12} = \frac{B^2}{2} + \tfrac{1}{6} \qquad (35)$$

as only q_2 and q_3 are correlated.

The successive values of q_i are assumed uncorrelated. These error steps $(q_i, q_{i+1} \ldots)$ are effectively added as noise $q(t)$ at the process input, as shown in Fig. 3(b).

The effect upon process output $c(t)$ of a step disturbance $q(t) = q_i(t - t_i)$ applied at the process input, is:

$$c_q(t) = q_i.\ h(t - t_i)$$

where $h(t)$ is illustrated in Fig. 4(c) for a typical controlled process.

The effect of the series of such disturbances is:

$$c_q(t) = \sum_{i=-\infty}^{j} q_i . h(t - iT) \qquad (36)$$

where

$$jT \leqslant t < (j+1)T.$$

The variance of $c_q(t)$, when all q_i are independent and of zero mean, is

$$\sigma_q^2 = < c_q^2(t) > \doteqdot < c_q^2(jT) >$$

so

$$\sigma_q^2 = < \sum_{i=-\infty}^{j} [q_i . h(t - iT)]^2 > \doteqdot < q_i^2 > \sum_{i=0}^{\infty} h^2(iT). \qquad (37)$$

If normalised sampling rate, V, is high, so that $h(t)$ can be described by samples at intervals of T, then

$$\sum_{i=0}^{\infty} h^2(iT) = \frac{1}{T}\int_0^{\infty} h^2(t) . \mathrm{d}t = \frac{\tau}{T} . k_4^2 . H \qquad (38)$$

where

$$H = \int_0^{\infty} h^2(t)\mathrm{d}t/\tau k_4^2$$

depends upon the type of process and the optimality of control. Typically $H = 0{\cdot}1$. Substituting (35) and (38) in (37) gives the variance of errors in output $c(t)$ due to quantisation:

$$\sigma_q^2 = (B^2/2 + \tfrac{1}{6})H.V.k_4^2$$

and assuming $B = B_{\mathrm{opt}}$, equation (31),

$$\sigma_q^2 = \underbrace{\left(\frac{3L_2}{k_2}\right)^2 \frac{H.V.}{2}}_{\text{input quantisation effects}} + \underbrace{k_4^2 . \frac{H.V.}{6}}_{\text{output quantisation effects}} \qquad (39)$$

As expected, the variance of the output $c(t)$ solely due to quantisation effects is reduced by using a longer signal register (k_2 increased) or finer motor steps (k_4 decreased). It is of interest to observe how this variance increases with V, the normalised sampling rate. For rapid sampling, quantisation errors exceed those due to other causes. Figure 5 shows a series of step responses for which V is varied and all other parameters held constant.

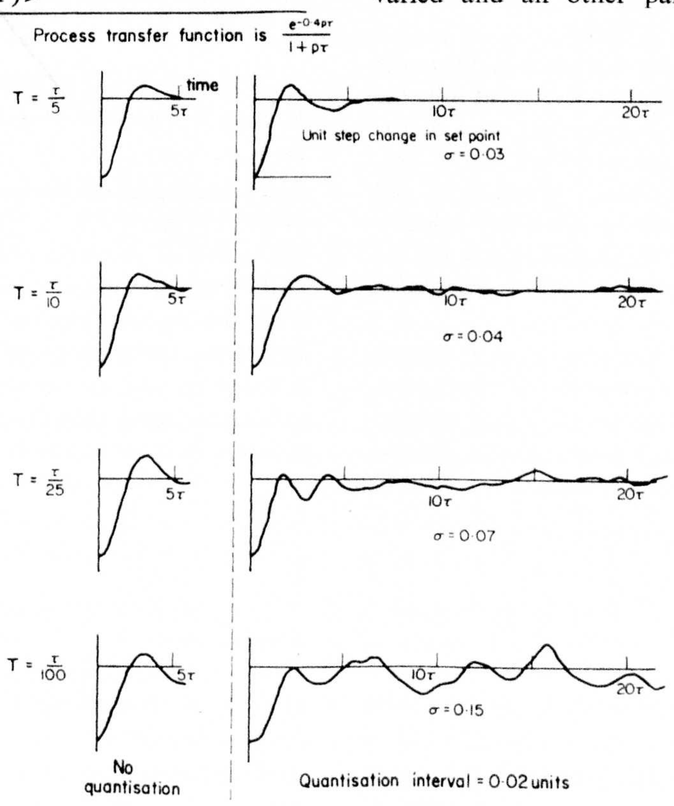


FIG. 5. Influence of sampling rate upon quantisation "noise".

8. EFFECT OF RUN-OUT TIME

No general theory has been developed to describe the effects of run-out time on stability. Simplified arguments suggest an increase in stability due to the longer sampling period, offset at very low sampling rates by the extra delays $u_a/2$ and $u_b/2$ introduced. A number of particular processes have been investigated, and in each case the run-out time tended to enhance stability. Figure 6 shows a typical result, except that the effects shown there reflect stepping rates for slower than those normally used. For control conditions well removed from large set point changes, the run-out time would rarely comprise more than 5% of the cycle time.

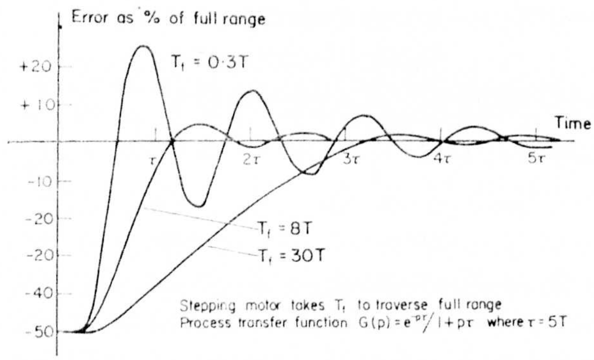


FIG. 6. Step responses with varying run-out time.

Work on similar systems [5] has confirmed the stabilising tendancy of run-out time in other particular cases.

9. CONCLUSIONS

The single loop controller described in this paper was developed as a digital replacement for the widely used analogue $P+I$ controller. As a digital device, it is free from parameter drift, easily programmed (e.g. for use in batch production) and capable of integration into a computer scheme. The frequency modulated transducer signal and the incremental output signal have good transmission properties.

Provided that the sampling rate lies in a suitable band, the controller performs like a continuous device; parameters can then be calculated without recourse to sampled data theory. V was defined as τ/T, the sampling rate normalised with respect to the dominant time constant of a process. For V less than 6, sampling effects become apparent and the control attainable deteriorates; for high values of V, quantisation effects degrade performance. Under certain conditions, these effects cause a variance of the controlled variable that is proportional to V. A digital controller designed for a wide range of process time constants thus requires either an adjustable sampling rate, or very fine input and output quantisation. Frequency counting methods introduce somewhat higher quantisation errors than conventional analogue-to-digital conversion.

Only the integral action of the controller is based upon the loop error; proportional action does not result directly from a change in set point. This arrangement is not usually advantageous and could be avoided by the extra complexity of defining the set point as a pulse rate. The interruption of the control cycle, while the stepping motor is moved, the run-out, has little effect upon normal performance, does not prejudice stability and tends to reduce the rate of change of loop variables during transients.

Acknowledgement—The authors wish to thank Serck R. and D., Birmingham for the assistance of their staff, and for permission to publish this paper. The digital controller described is a Serck invention (patent applied for) and this analysis was the subject of a research contract.

REFERENCES

[1] G. J. MOSHOS: Survey of countup-countdown machines. *IFAC Pulse Symp.*, Budapest, April (1968), Paper A23.
[2] J. L. DOUCE, K. C. NG and K. K. MURTHY: How the frequency response of a sample-and-hold unit should be expressed. *Control* **11**, No. 112, pp. 492–494, October (1967).
[3] A. J. MONROE: *Digital Processes for Sampled-data Systems*, Chap. 21. Wiley, New York (1962).
[4] B. WIDROW: A study of rough amplitude quantisation by means of Nyquist sampling theory. *IRE Trans. Circuit Theory* **CT-3**, 266–276 (1956).
[5] G. B. ASTARDJIAN, P. A. PETROV and H. R. HRISTOV: On a principle of sampled-data control systems with pulse-width modulation and variable signal-dependent sampling. IFAC, London, June 1966, Paper 34A.
[6] J. M. NIGHTINGALE and G. A. RICHARDS: Error analysis in binary rate modulation systems. *IFAC Pulse Symp.*, Budapest, April (1968), Paper A21.

APPENDIX

Statistical properties of quantisation errors

WIDROW [4] has considered the conditions under which the statistical properties of the errors introduced by quantising a signal are independent of the statistical properties of the signal. He showed that in general the error introduced is more random than the signal. He showed in particular, that where the probability density of a signal entering a quantiser can be reconstructed from that of the quantised signal, then the probability density of the introduced error is a box function extending from $-\frac{1}{2}$ to $+\frac{1}{2}$ of a quantisation level. He further indicated that the errors introduced by quantising successive samples are almost uncorrelated, if the samples are separated by several levels.

The quantisation associated with frequency-to-number conversion is a more complex phenomenon than conventional amplitude quantisation. NIGHTINGALE and RICHARDS [6] have developed expressions for the error introduced by the entire process of frequency encoding and decoding. A more limited analysis is adequate here.

Consider a pulse train whose frequency f varies over a limited band, so that f has a standard deviation small compared with its mean. By counting over a gate period, of length u, frequency f is converted to a number in the series M_i, M_{i+1} etc.

$$M_i = m_i + h_i = \int_{(i-1)u}^{i.u} f . \mathrm{d}t$$

where m_i is integer, and $0 \leqslant h_i < 1$.

The number observed in the register, after the ith frequency-to-number conversion, will be either m_i, or $(m_i + 1)$, giving quantisation errors of $q_i = -h_i$ and $q_i = 1 - h_i$ respectively.

If the series M_i, M_{i+1} etc. ranges over several numbers, and if $m_{i+1} \sim m_i$ is usually >2 (say), then the series h_i, h_{i+1} . . . will have statistical properties independent of the M_i, M_{i+1}, . . . series.

WIDROW considered the normal rounding.

$$q_i = h_i \text{ for } h_i < \tfrac{1}{2}\,; \qquad q_i = 1 - h_i \text{ for } h_i \geqslant \tfrac{1}{2}$$

For frequency counting, however, q_i may equal either h_i or $1 - h_i$ with a known probability attached to each. This uncertainty is caused by variations in the *phase* of the frequency signal at the instant of opening the counting gate. Moreover, the phase at the start of one conversion is determined by the phase and also (h_i) for the previous conversion.

Table 2 summarises the differences between frequency-to-number conversion, and conventional quantisation. Note how frequency counting methods introduce the larger quantisation errors, which for adjacent samples are correlated.

TABLE 2. STATISTICAL PROPERTIES OF QUANTISATION ERRORS WHEN INPUT TO QUANTISER SATISFIES WIDROW'S CRITERIA
(Calculated values confirmed by experiment)

Conventional A/D convertor	Frequency-to-number convertor
Probability density	
$p(q)=0 \quad q<-\frac{1}{2}, \quad q\geqslant\frac{1}{2}$	$p(q)=0 \quad q<-1, \quad q\geqslant 1$
$p(q)=1 \quad -\frac{1}{2}\leqslant q<\frac{1}{2}$	$p(q)=(1+q)/2 \quad -1\leqslant q<0$
	$p(q)=(1-q)/2 \quad 0\leqslant q<1$
Expected value $<q>=0$	$<q>=0$
Variance $<q^2>=1/12$	$<q^2>=1/6$
Covariance	
$<q_iq_{i+n}>=0 \quad n>0$	$<q_iq_{i+n}>=1/12 \quad n=1$
	$=0 \quad n>1$
Variance of error of difference	
$<(q_{i+1}-q_i)^2>=1/6$	$<(q_{i+1}-q_i)^2>=1/2$

Résumé—Les techniques numériques dans la commande des processus trouvent leurs applications à plusieurs niveaux de complexité. En particulier, il y a de fortes tendances à considérer l'emploi de petits régulateurs numériques de boucle unique pour remplacer leur contrepartre analogique plus conventionnelle. Cet article présente les résultats d'une étude combinée, theorique et expérimentale, d'un régulateur numérique à deux actions, qui fonctionne sur la sortie d'un transmitteur de mesure à fréquence variable d'impulsions et qui émet des trains d'impulsions pour régler une variable du processus par l'intermédiaire d'un moteur pas-à-pas.

La caractéristique distinctive d'un tel système est que les considérations économiques exigent des algorithmss de commande relativement simples. Ceux-ci introduisent un certain nombre de non-linéarités dans les équations d' ensemble de la commande. L'article considère les conditions dans lequelles ces non-linéarités et les effets d'échantillonnage sont insignifiants, de sorte que le régulateur se comporte effectivement comme un dispositif continu. Certains de ces effets non-linéares sont analysés et il est montre que, lorsque'ils ne sont pas négligeables, ils tendent à dégrader les performances du régulateur. Un paramètre V, le rapport de la constante de temps prédominante du processus à la période d'échantillonnage, est utilisé pour définir une région de performances satisfaisantes.

Zusammenfassung—Die Digitaltechnik wird bei der Prozeßkontrolle in verschiedenen Schwierigkeitsgraden angewandt. Speziell liegt ein starker Anreiz zur Betrachtung kleiner einschleifiger digitaler Regelkreise vor, um durch sie ihre konventionelleren analogen Gegenstücke zu ersetzen. Diese Arbeit liefert die Ergebnisse einer kombinierten theoretischen und experimentellen Studie eines digitalen Regelkreises, der auf den Ausgang eines Meßwertwandlers, der eine variable Pulsfrequenz mißt, arbeitet, und an seinem Ausgang einen Impulszug erzeugt, um über den Schrittmotor eine Prozeßvariable zu justieren.

Das kennzeichnende Merkmal eines solchen Systems ist, daß ökonomische Betrachtungen relativ einfache Kontrollalgorithmen erfordern. Diese schleppen in die Gleichungen eine Anzahl von Nichtlinearitäten ein. Betrachtet werden die Bedingungen, unter denen diese Nichtlinearitäten und die Abtasteffekte geringfügig sind, so daß sich der Regelkreis tatsächlich als kontinuierliche Einrichtung verhält. Einige der nichtlinearen Effekte werden analysiert und es wird gezeigt, daß sie, wenn sie merklich werden, zu einer Herabsetzung der Leistungsfähigkeit tendieren. Ein Parameter V, das Verhältnis der dominierenden Prozeßzeitkonstanten zur Abstastperiode, wurde benutzt, um einen Bereich befriedigender Arbeitsweise zu definieren.

Резюме—Цифровые техники в управлении процессами находят применения на различных уровнях сложности. В частности, имеется сильная тенденция рассматривать применение небольших цифровых результатов единственного контура для замены более классических аналоговых регуляторов. Настоящая статья предлагает результаты смешанного теоретического и экспериментального исследования цифрового регулятора с двумя воздействиями, работающего на выходе измерительного датчика переменной частоты импульсов и выдающего серии импульсов для регулирования переменной процесса путем шагового двигателя.

Отличной характеристикой такой системы является факт что экономические соображения требуют сравнительно простых управляющих алгоритмов, последние вводят некоторое число нелинейностей в уравнения цельности управления. Статья рассматривает условия в которых этими нелинейностями и влиянием квантования можно пренебречь, так что регулятор ведет себя я самом деле как аппарат непрерывного действия. Некоторые из этих нелинейных эффектов проанализированы и показано что, когда ими нельзя пренебречь, они стремятся ухудить работу регулятора. Используется параметр V, отношение главной временной постоянной процесса к периоду квантования, для определения зоны удовлетворительной работы.